

General Disclaimer

One or more of the Following Statements may affect this Document

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

Accuracy of Remotely Sensed Data
Sampling and Analysis Procedures

Virginia Polytechnic Inst. and State Univ.
Blacksburg

Prepared for

Forest Service
Houston, TX

Jan 82



AgRISTARS

RR-U2-04257

A Joint Program for Agriculture
and Resources Inventory
Surveys Through Aerospace
Remote Sensing

January 1982

NATIONWIDE FORESTRY APPLICATIONS PROGRAM Renewable Resources Inventory Project

Cooperative Research Report NFAP-276

ACCURACY OF REMOTELY SENSED DATA: SAMPLING AND ANALYSIS PROCEDURES

Cooperative Agreement No. 13-1134

Virginia Polytechnic Institute
and State University

Blacksburg, Virginia 24061



USDA Forest Service



REPORT DOCUMENTATION PAGE		1. REPORT NO. NFAP-276	2.	3. Recipient's Accession No. PBB-2 25043 2
4. Title and Subtitle Accuracy of Remotely Sensed Data: Sampling and Analysis Procedures				
7. Author(s) R. G. Congalton; R. G. Oderwald; and R. A. Mead				
9. Performing Organization Name and Address School of Forestry and Wildlife Resources VPI & State University Blacksburg, VA 24061				
12. Sponsoring Organization Name and Address USDA, Forest Service NF Program 1050 Bay Area Blvd. Houston, TX 77058				
15. Supplementary Notes				
16. Abstract (Limit: 200 words) The main body of this report is divided into two parts. The first part presents a review and update of the discrete multivariate analysis techniques used for accuracy assessment. Appendix A contains a listing of the computer program written to implement these techniques. The second part presents new work on evaluating accuracy assessment using Monte Carlo simulation with different sampling schemes.				
Appendix B contains the results of the accuracy assessment analysis for the eight error matrices from the mapping effort of the San Juan National Forest. Appendix C contains a method of estimating the sample size requirements for implementing the accuracy assessment procedures. Appendix D contains a proposed method for determining the reliability of change detection between two maps of the same area produced at different times.				
17. Document Analysis a. Descriptors Accuracy assessment Change detection Remote sensing San Juan National Forest				
b. Identifiers/Open-Ended Terms				
c. COSATI Field/Group Forestry (remote sensing)				
18. Availability Statement Release Unlimited				
19. Security Class (This Report) Unclassified			21. No. of Pages 93	
20. Security Class (This Page) Unclassified			22. Price	

Project Supported by

Nationwide Forestry Applications Program
Renewable Resources Inventory Project
Houston, Texas

Cooperative Agreement Number 13-1134

January 1982

Remote Sensing Research Report 82-1

Accuracy of Remotely Sensed Data:
Sampling and Analysis Procedures

by

Russell G. Congalton 1/
Richard G. Oderwald 2/
Roy A. Mead 3/

- 1/ Graduate Research Assistant, School of Forestry and Wildlife Resources, VPI & SU, Blacksburg, VA 24061.
- 2/ Associate Professor, School of Forestry and Wildlife Resources, VPI & SU, Blacksburg, VA 24061.
- 3/ Assistant Professor, School of Forestry and Wildlife Resources, VPI & SU, Blacksburg, VA 24061.

TABLE OF CONTENTS

List of Tables

List of Figures

Preface

Abstract

1.0 Introduction

2.0 Discrete Multivariate Analysis Techniques for Accuracy Assessment

 2.1 Review of the Normalization Procedure

 2.2 Update of the Test of Agreement Procedure

 2.3 Update of KAPPA Example

 2.4 Review of the Multi-factor Comparison Procedure

 2.5 Multi-factor Comparison Example

3.0 Sampling Simulation for Accuracy Assessment

 3.1 Objectives

 3.2 Study Areas

 3.2.1 Forest Land Environment

 3.2.2 Rangeland Environment

 3.2.3 Agricultural Land Environment

 3.3 Data

 3.4 Procedure

 3.5 Results

 3.5.1 Different Images

 3.5.2 Intra-cluster correlation coefficients

 3.6 Conclusions

 3.7 Further Work

4.0 Literature Cited

- Appendix A Updated Listing of FORTRAN Computer Program KAPPA
- Appendix B Results of Accuracy Assessment of the San Juan National Forest R2MAP Land Cover Classification
- Appendix C Estimating Sample Size Requirements for Accuracy Assessment using KAPPA
- Appendix D A Proposed Method for Estimating the Reliability of Change Detection
- Appendix E Accuracy Assessment of the San Juan National Forest R2MAP Land Cover Classification
- Appendix F Explanation of the Two Land Cover Classification Systems used in the San Juan National Forest Accuracy Assessment

List of Tables

<u>Table</u>	<u>Title</u>	<u>Page</u>
1.	Table of updated KHAT values	6
2.	A list of factors for the four-way table comparing enhancement and classification algorithms	10
3.	The original four-way table for comparing enhancement techniques and classification algorithms.	11-14
4.	A list of the land cover categories for the four-way table comparing enhancement and classification methods	15
5.	The uniform order models for the four-way table comparing enhancement techniques and classification algorithms	16
6.	The model selection process for the four-way table comparing enhancement techniques and classification algorithms	17-20
B1.	Kappa (κ), variance of Kappa, and 95% confidence intervals for Kappa for each classification matrix	40
B2.	Tests for significant differences between Kappa's from selected matrices	42

List of Tables

(Continued)

<u>Table</u>	<u>Title</u>	<u>Page</u>
E1.	The contract land cover categories and the corresponding R2MAP symbols	52
E2.	The data base land cover categories and the corresponding R2MAP symbols	53-54
E.3	Contract Classification - East Half	58
E.4	Contract Classification - West Half	59
E.5	Abbreviated Contract Classification - East Half	60
E.6	Abbreviated Contract Classification - West Half	61
E.7	Data Base Classification - East Half	62
E.8	Data Base Classification - West Half	63
E.9	Abbreviated Data Base Classification - East Half	64
E.10	Abbreviated Data Base Classification - West Half	65

List of Figures

<u>Figure</u>	<u>Title</u>	<u>Page</u>
1.	Flow diagram of sampling simulation procedure	22
2.	Difference image for agricultural environment	27
3.	Difference image for range environment	28
4.	Difference image for forest environment	29
5.	A plot of average ROH vs. cluster size for each vegetation environment	30
B1.	Graphical representation of 95% confidence intervals for Kappa for each classification matrix	41
D1.	Example of the proposed methods of determining reliability of change detection	50

PREFACE

The authors would like to thank the Applications Branch of the EROS Data Center for their help in the sampling simulation section of this report. Special thanks go to Tom Loveland and Wayne Miller for supplying the data and also to John Szaigin and Jeff Newcomer for their guidance and encouragement.

Acknowledgment should also go to Robert B. Mattson, James A. Shafer, and Bruce F. Short for their work or the photo interpretation used in the accuracy assessment of the R2MAP classification for the San Juan National Forest. Thanks also to all those who helped code all the data and prepare it for analysis.

ABSTRACT

The main body of this report is divided into two parts. The first part presents a review and update of the discrete multivariate analysis techniques used for accuracy assessment. Appendix A contains a listing of the computer program written to implement these techniques. The second part presents new work on evaluating accuracy assessment using Monte Carlo simulation with different sampling schemes.

Appendix B contains the results of the accuracy assessment analysis for the eight error matrices from the mapping effort of the San Juan National Forest. Appendix C contains a method of estimating the sample size requirements for implementing the accuracy assessment procedures. Appendix D contains a proposed method for determining the reliability of change detection between two maps of the same area produced at different times.

1.0 Introduction

This report is divided into two parts. The first part deals with a short review and update of material described in last year's report (Congalton et al. 1981). This work involves assessing the accuracy of remotely sensed data using discrete multivariate analysis statistical techniques.

The second part of this report describes the work currently in progress on sampling for accuracy assessment. This research is investigating different sampling schemes using Monte Carlo simulation techniques. Although this work is not complete, some valuable results have already been achieved.

2.0 Discrete Multivariate Analysis Techniques for Accuracy Assessment

The three analysis procedures reviewed here all involve error matrices. An error matrix is a square array of numbers set out in rows and columns which express the number of cells assigned as a particular land cover type relative to the actual cover type as verified in the field. The columns usually represent the reference data and the rows indicate either the Landsat classification or the photo interpretation. The discrete multivariate analysis procedures are performed on the error matrices.

2.1 Review of the Normalization Procedure

The first comparison procedure (Bishop *et al.* 1975) allows individual cell values in each error matrix to be compared. This comparison is made possible by a process called normalizing the error matrix. This normalization process is a way of standardizing each matrix so that a direct comparison of individual cell values is possible. This procedure always converges to a unique set of maximum likelihood estimates and as such is the best algorithm to use in this case (Fienberg 1970). An assumption made by this process is that all cells are of equal importance.

Normalization of an error matrix is an iterative process by which the rows and columns of the matrix are successively balanced until each row and column adds up to a given value (marginal). This process causes each cell value to be influenced by all the other cell values in its corresponding row and column. Each cell value is then a combination of reference data and remote sensor data and is representative of both commission and omission errors for that land cover category. Because each row and column must add to a given marginal, the cell values in corresponding positions of two or more error matrices can then be compared without regard for differences in sample size between matrices.

The normalization process is performed by a computer program called MARGFIT (Congalton *et al.* 1981). For additional details and examples of this process see Congalton (1981).

2.2 Update of the Test of Agreement Procedure

The second method of comparison is a procedure that tests for agreement between two or more error matrices (Bishop *et al.* 1975). This measure of agreement is based on the difference between the actual agreement of the classification (i.e., agreement between remote sensor data and reference data indicated by the major diagonal) and the chance agreement which is indicated by the row and column marginals. This measure of agreement called KHAT is calculated by:

$$\hat{k} = \frac{\frac{\sum_{i=1}^n x_{ii} - \sum_{i=1}^n (x_{i+} * x_{+i})}{n^2 - \sum_{i=1}^n (x_{i+} * x_{+i})}}$$

where:

n is the number of rows in the matrix

x_{ii} is the number of observations in row i and column i

x_{i+} and x_{+i} is the marginal total of row i and column i respectively
and N is the total number of observations.

A KHAT value is calculated for each matrix and is a measure of how well the Landsat classification or photo interpretation agrees with the reference data. The approximate large sample variance of KHAT as determined by the delta method is:

$$\hat{\sigma}(\hat{k}) = \frac{1}{N} \frac{\theta_1(1 - \theta_1)}{(1 - \theta_2)^2} + \frac{2(1 - \theta_1)(2\theta_1\theta_2 - \theta_3)}{(1 - \theta_2)^3} + \frac{(1 - \theta_1)^2(\theta_4 - 4\theta_2^2)}{(1 - \theta_2)^4}$$

where:

$$\theta_1 = \sum_{i=1}^n \frac{x_{ii}}{N} \quad \theta_3 = \sum_{i=1}^n \frac{x_{ii}}{N} \left(\frac{x_{i+}}{N} + \frac{x_{+i}}{N} \right)$$

$$\theta_2 = \sum_{i=1}^n \frac{x_{i+} * x_{+i}}{N^2} \quad \theta_4 = \sum_{i=1}^n \sum_{j=1}^n \frac{x_{ij}}{N} \left(\frac{x_{+j}}{N} + \frac{x_{i+}}{N} \right)^2$$

Confidence intervals can be calculated for KHAT using this approximate large sample variance of KHAT. These confidence intervals were used previously as a method for testing the significant difference between two error matrices. However, exact hypothesis tests are now available and should be used instead of the confidence intervals.

A test for significance of KHAT can be performed to determine if the agreement between the Landsat classification or photo interpretation and reference data is significantly greater than zero. Also a test for the significant difference between two independent KHAT's can be performed by evaluating the normal curve deviate (Cohen 1960). The test statistic for significant difference is approximately:

$$\sqrt{\frac{\hat{k}_1 - \hat{k}_2}{\hat{\sigma}_1^2 + \hat{\sigma}_2^2}} \sim z$$

The FORTRAN computer program used to calculate this measure of agreement, KHAT, is called KAPPA. This program has been updated to

include the exact hypothesis tests described above (Appendix A). Given the original matrix, the computer program implements a procedure that calculates the KHAT value and its corresponding variance. A confidence interval around KHAT is also computed along with the test statistic for significance of KHAT. All these values plus the values used in calculating the variance (i.e., TH1, TH2, TH3, and TH4) are printed out along with the original error matrix. The algorithm then computes the test statistic for significant difference between independent KHAT's for each possible pair of matrices. These values are printed out in a summary table at the end of the program.

2.3 Update of KAPPA Example

As already mentioned, an actual test statistic is now available to test for significant differences between error matrices. In last year's report (Congalton et al. 1981) examples were given in which only the confidence intervals were compared. Presented below is an updated analysis of the results that appeared last year in Table 6, page 19. This data compares four classification algorithms provided by Hoffer (1975).

Table 1. Table of updated KHAT values.

MATRIX	KHAT	VARIANCE	COMPARISON	Z STATISTIC	Result	
					95%	90%
Non-supervised (10 cluster) NS-10	0.60479	.00073735	NS-10, NS-20	0.47475	NS	NS
Non-supervised (20 cluster) NS-20	0.58573	.00087456	NS-10, MS	3.00930	S	S
			NS-10, MC	-2.93550	S	S
Modified Supervised MS	0.47581	.00109972	NS-20, MS	2.47390	S	S
			NS-20, MC	-3.28090	S	S
Modified Clustering MC	0.71846	.00076218	MS, MC	-5.62360	S	S

This analysis shows that there is not a significant difference between the classification obtained using a nonsupervised approach with 20 clusters and that obtained using a nonsupervised approach with 10 clusters. However, the results of all the other tests yield significant differences between classification algorithms.

2.4 Review of the Multi-factor Comparison Procedure

The multi-factor comparison procedure allows more than one factor affecting the classification accuracy to be examined at the same time. The log-linear approach as described by Fienberg (1980) and Bishop et al. (1975) is a method by which many variables and the interaction between these variables can be tested simultaneously to see which are necessary (i.e., significant) for explaining the classification accuracy.

The simplest model (combination of variables) that provides a good fit to the data is chosen using a model selection procedure. This procedure allows the user to systematically search all possible models and choose the simplest model that provides a good fit to the data. First all uniform order models are tested (i.e., models with all possible n-way interactions, where n ranges from 1 to the number of factors) and the simplest good fit model is chosen. Each interaction of the chosen model is then tested for significance. If the interaction is not significant it is dropped until a model is found in which all the factors and

interactions of factors are significant. For a more detailed description of this stepwise model selection procedure, see Fienberg (1980) Section 5.3. The criteria for selecting a good model is based on a Likelihood Ratio, G^2 , and the degrees of freedom for the model. The Likelihood Ratio has an asymptotically chi-square distribution and therefore the critical value for testing if the model is a good fit can be obtained from a chi-square table using the appropriate degrees of freedom.

The Likelihood Ratio is calculated using an Iterative Proportional Fitting procedure (Fienberg 1980 and Bishop et al. 1975). This procedure uses a method of successive approximations to converge to the maximum likelihood estimates of the minimum sufficient statistics as defined by the model. Therefore, the log-linear approach allows for analysis of multi-way tables with many factors. For example, error matrices generated using different dates, different algorithms, and different analysts all of the same scene of imagery can be put together and the factors necessary to explain the classification accuracy analyzed. This example would yield a five-way table with the five factors being: date, algorithm, analyst, Landsat classification, and reference data. Testing this five-way table would determine the simplest model of factors and interactions that best explain the results.

2.5 Multi-factor Comparison Example

The data used to test the combined effects of different classification algorithms and enhancement techniques on Landsat classification accuracy was supplied by Gregg et al. (1979). In this example two classification algorithms are performed on smoothed and unsmoothed imagery and the combined effects are studied. The factors and effects for this four-way table are listed in Table 2 and the original matrices presented in Table 3. Each algorithm classified the data into one of ten land cover categories (Table 4).

A model selection procedure was performed on the four-way table beginning with the uniform order models (Table 5). The results of this procedure yields the simplest best fit model to the data (Table 6). This model, [14] [24] [34], indicates that no three or four-way interactions are needed to explain the data. Instead, there are only two-way interactions involved. In other words, there is a combined effect due to each explanatory variable (i.e., algorithm, enhancement, and reference data) separately with the response variable. However, there are no higher order interactions. Therefore, each effect is important and no factor can be eliminated. The assumption that the error matrices adequately represent the actual classification must hold here if any of these results are to be meaningful.

TABLE 2

A list of factors for the four-way table comparing enhancement and classification algorithms.

FACTOR	EFFECT	
I (1)	Classification Algorithm	1=maximum likelihood 2=cononical domain
J (2)	Resampling Technique	1=smoothed 2=unsmoothed
K (3)	Reference Data	1-10 (See Table 4)
L (4)	Landsat Data	1-10 (See Table 4)

TABLE 3

The original four-way table for comparing enhancement techniques and classification algorithms.

		reference data (X)									
		1	2	3	4	5	6	7	8	9	10
land data (Y)	1	69	128	0	23	4	0	2	4	9	0
	2	17	60	0	9	0	0	0	7	15	2
	3	13	38	2	14	1	0	2	5	10	1
	4	16	57	3	165	31	4	7	7	17	5
	5	0	5	6	38	33	17	12	7	3	17
	6	17	19	0	3	2	0	1	16	76	3
	7	2	0	1	1	9	7	27	9	0	5
	8	3	23	4	22	8	14	10	164	173	5
	9	5	6	1	11	7	0	5	113	569	1
	10	0	0	3	6	4	4	3	0	0	15

*See Table 4 for a list of the land cover categories.

TABLE 3
(continued)

		reference data (X)									
		1	2	3	4	5	6	7	8	9	10
1	37	166	0	29	6	1	3	6	20	0	
2	16	50	0	10	1	1	0	6	13	1	
3	5	23	1	13	1	0	0	2	6	1	
4	12	49	3	16 ^a	26	2	4	7	13	2	
land use (L)	5	1	0	6	29	39	15	8	10	9	21
	6	7	11	0	3	1	1	1	9	45	2
	7	3	0	3	0	6	7	33	8	1	11
	8	5	22	2	21	8	19	12	174	159	4
	9	11	9	0	12	3	3	2	110	617	1
	10	0	0	7	2	3	3	3	0	0	11

^aSee Table 4 for a list of the land cover categories.

TABLE 3
(continued)

		reference data (X)									
		1	2	3	4	5	6	7	8	9	10
1	59	112	0	29	3	0	1	4	3	0	0
2	22	50	0	10	1	0	0	6	16	1	
3	14	60	1	23	1	2	2	7	13	1	
4	15	55	3	129	25	4	4	6	12	0	
5	0	0	3	28	30	15	12	3	5	10	
6	11	15	1	5	2	1	0	9	47	3	
7	0	0	1	1	3	6	28	9	3	7	
8	0	27	0	24	11	17	11	156	157	6	
9	16	11	1	13	6	3	6	132	626	4	
10	0	0	6	2	2	3	2	0	0	10	

*See Table 4 for a list of the land cover categories.

TABLE 3
(continued)

		reference data (%)									
		1	2	3	4	5	6	7	8	9	10
Landunit data (1)	1	70	141	0	28	-	1	2	6	19	0
	2	26	53	0	6	1	1	0	4	10	0
	3	11	48	0	25	3	0	0	4	8	1
	4	12	53	6	167	34	1	4	6	17	2
	5	0	1	10	25	00	15	6	8	8	22
	6	1	5	1	5	1	1	0	4	18	0
	7	2	0	0	0	5	3	34	3	0	14
	8	4	20	2	12	9	21	14	164	127	5
	9	21	15	0	17	8	3	0	128	681	2
	10	0	0	5	1	1	1	0	0	0	3

*See Table 4 for a list of the land cover categories.

TABLE 4

A list of the land cover categories for the four-way table comparing enhancement and classification methods.

CATEGORY NUMBER	LAND COVER TYPE
#1	disturbed/recent clearcut
#2	planted clearcut
#3	mixed reproduction hardwood/conifer
#4	young reproduction (conifer)
#5	old reproduction (conifer)
#6	mixed pole/saw timber (hardwood/conifer)
#7	conifer - pole timber
#8	conifer - saw timber
#9	old growth
#10	hardwood

TABLE 4

A list of the land cover categories for the four-way table comparing enhancement and classification methods.

CATEGORY NUMBER	LAND COVER TYPE
#1	disturbed/recent clearcut
#2	planted clearcut
#3	mixed reproduction (hardwood/conifer)
#4	young reproduction (conifer)
#5	old reproduction (conifer)
#6	mixed pole/saw timber (hardwood/conifer)
#7	conifer - pole timber
#8	conifer - saw timber
#9	old growth
#10	hardwood

TABLE 5

The uniform order models for the four-way table comparing enhancement techniques and classification algorithms.

MODEL	χ^2	df	RESULT
{1}{2}{3}{4}	10000.07200	152	poor fit
A	165.06620	29	good fit
{12}{13}{14}{23}{24}	20.90917	9	good fit

TABLE 6

The model selection process for the four-way table comparing enhancement techniques and classification algorithms.

MODEL	χ^2	df	RESULT
(12)(13)(14)(23)(24)	10732.65762	315	poor fit
(12)(13)(14)(23)(36)	230.22106	263	good fit
(12)(13)(14)(24)(36)	147.82266	263	good fit
(12)(13)(23)(24)(36)	227.51772	263	good fit
(12)(14)(23)(24)(36)	156.51131	263	good fit
(13)(14)(23)(24)(36)	146.06061	235	good fit

model B best and good fit

$\chi^2(B) = \chi^2(A) = 2.06619 \sim \chi^2_{9df}$ not significant so drop [23]

TABLE 6
(continued)

MODEL.	χ^2	df	RESULT
[12][13][14][2b]	1073.0270	324	poor fit
[12][13][14][3b]	231.3918	252	good fit
[12][13][2b][3b]	229.4002	252	good fit
[12][1b][2b][3b]	158.4941	252	good fit
[13][2b][3b] C	160.0036	252	good fit

model C has a good fit

$\chi^2(C) - \chi^2(B) = 0.1709195 - \chi^2_{1df}$ not significant no drop [12]

TABLE 6
(continued)

MODEL	χ^2	$\Delta\chi^2$	RESULT
[13][14][15]	10733.99876	325	poor fit
[14][15][16]	231.63685	253	good fit
[15][16]	229.52108	253	good fit
[14][26][16]	150.66500	253	good fit

model D best and good fit

$\chi^2(D) - \chi^2(C) = 10.66162 \sim \chi^2_{1df}$ not a significant drop [13]

TABLE 6
(continued)

MODEL.	α^2	df	RESULT
[14][24][3]	10736.29202	363	poor fit
[14][34][2]	262.09646	262	good fit
[24][34][1] E	229.01433	262	good fit

model E has a good fit
 $\alpha^2(E) - \alpha^2(D) = 71.1699 \sim \chi^2_{9df}$ not significant so
 can't drop [14]

3.0 Sampling Simulation for Accuracy Assessment

This research involved a sampling simulation study using three different vegetation environments of varying spatial complexity. These three environments were forest, range, and agricultural lands. Small areas (approximately 200 x 200 pixels) called subscenes were chosen from each of the three environments. Some of these subscenes contained large homogeneous areas of vegetation while others had very diverse vegetation. Associated with each subscene were two classified data sets which were compared with each other to create a difference image. A difference image is a matrix of zeros and ones, where the zeros indicate agreement between the two data sets and the ones indicate disagreement. The population parameters were computed from a 100% sample (i.e., total enumeration) of the difference image. The difference image was also repeatedly sampled with various sampling schemes using Monte Carlo methods. A flow diagram of this procedure is displayed in Figure 1.

3.1 Objectives

The objectives of this research were to determine the best (minimum variance) unbiased sampling method to use on a given vegetation environment. This vegetation environment was then related to a pattern of classification error. Once the pattern of error was known, it was then possible to relate this sampling method to other areas of similar patterns of classification error.

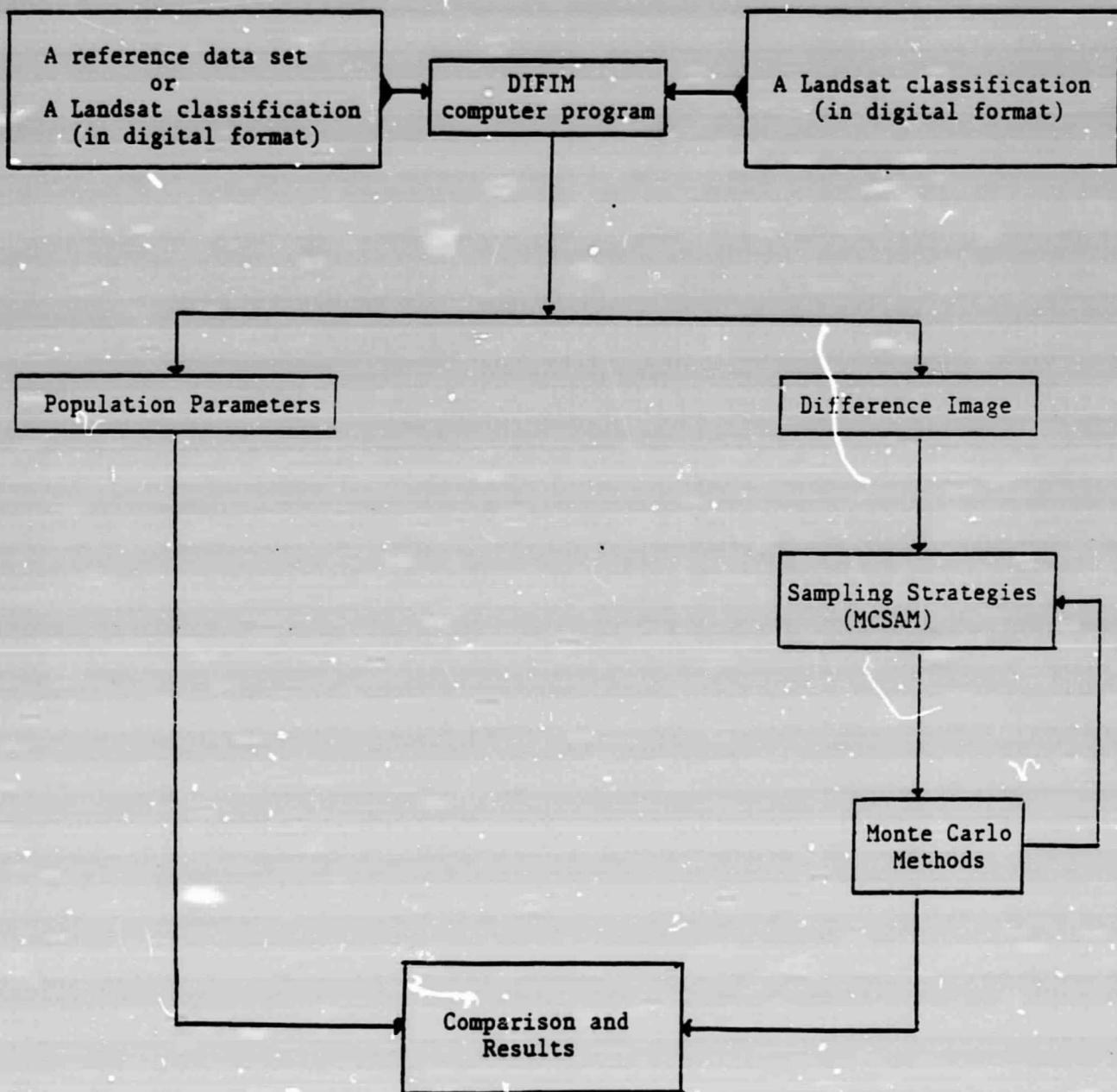


Figure 1. Flow diagram of sampling simulation procedure.

3.2 Study Areas

3.2.1 Forest Land Environment

The forest land study area that was used in this project is the Lolo Creek area located in western Montana. The Bitterroot mountains are the dominant physical feature of the area with elevations ranging from 3,000 feet to 9,200 feet. Average precipitation varies between 50 and 66 centimeters per year. The vegetation of the area is characterized by intermountain forest species.

The subscene chosen for use in this project was the Garden Point $7\frac{1}{2}$ minute quadrangle. This subscene was classified using a 60 meter pixel and resulted in 12 land cover categories. Four of the land cover categories were roads while the other eight were vegetation types.

3.2.2 Rangeland Environment

The rangeland study that was used in this project is located in the northwest corner of Arizona in Mojave County. The area is approximately 1,000,000 hectares in size and is representative of a southwestern desert environment. The Colorado River is the major drainage for the region. The area has a climate characterized by light precipitation, moderate temperatures, plentiful sunshine and low humidity. The vegetation varies from creosote bush and blackbrush at lower elevations to pinyon-juniper and ponderosa pine at higher elevations. The rangeland subscene chosen out of this study area was the Lizard Point $7\frac{1}{2}$ minute quadrangle. This subscene was classified into nine land cover categories. The pixel size used here was 50 meters.

3.2.3 Agricultural Land Environment

The agricultural land study area that was used in this project is the Umatilla Basin which occupies approximately 1.6 million acres in northcentral Oregon. This region is bounded to the north by the Columbia River. The area is characterized by an arid climate averaging less than 10 inches of precipitation per year.

Center pivot irrigation is the major type of irrigation used in the northern section of the basin. It is in this area that a subscene was taken for study in this project. Data from the Clarke 7½ minute quadrangle was available in Landsat classification and digitized reference data form. The classification was performed using a pixel size of one acre.

3.3 Data

At least one Landsat classification was available for each subscene. For the forest and range study areas two Landsat classifications were available. The forest study area had one classification performed using DMA terrain data with the Landsat data while the other classification used DEM terrain data alongwith Landsat. The range study area had one classification performed using DMA terrain data along with the Landsat data while the other classification was based on the Landsat data alone.

The agriculture study area only had one Landsat classification. The other data set used was a reference data set derived from digitizing photography and land surveys.

3.4 Procedure

Once all the data was in digital format, a difference image was generated for each subscene using the computer program, DIFIM. This program processed the two corresponding data sets for each subscene pixel by pixel. When the two corresponding pixels were classified the same, a zero was stored in that place in the output image. If the two corresponding pixels were classified differently, then a one was stored in that place in the output image. Therefore, an output image of zeros and ones was created and called the difference image.

The difference image was then used to generate the population parameters for each subscene. Since the population (i.e., the subscene) was binomially distributed (i.e., a matrix of zeros and ones), the parameters of interest were the size of the population, N , the proportion of correct responses, P , and the variance. The population parameters were also calculated within the DIFIM program.

After these calculations were completed each subscene was repeatedly sampled using Monte Carlo methods and different sampling strategies. These sampling simulations were performed by a computer program called MCSAM. The required inputs for this program were the sampling scheme, the sample size, and the number of repetitions. The outputs of this program were the sample mean, sample variance, and the

number of times the population mean was not contained within the sample confidence interval. For cluster sampling the outputs also included a measure of relative efficiency and the intra-cluster correlation coefficient.

3.5 Results

As previously mentioned, all the results for this research have not been completed. Some preliminary results are given below.

3.5.1 Difference images

The difference images created for each vegetation environment are in Figures 2-4. Note that the yellow shows the areas of agreement between the two data sets while the blue represents pixels of disagreement. Also notice the patterns of error for each vegetation environment.

3.5.2 Intra-cluster correlation coefficients

When using cluster sampling the effects of the cluster need to be measured. A measure of the homogeneity of the cluster is called ROH, intra-cluster correlation. The more homogeneous a cluster the greater the value of ROH. Intuitively one would like the cluster to be as diverse (i.e., heterogeneous) as possible so as to gain maximum information. Therefore it is desirable for ROH to approach zero. Figure 5 shows a plot of average ROH vs. cluster size for each of the vegetation environments.

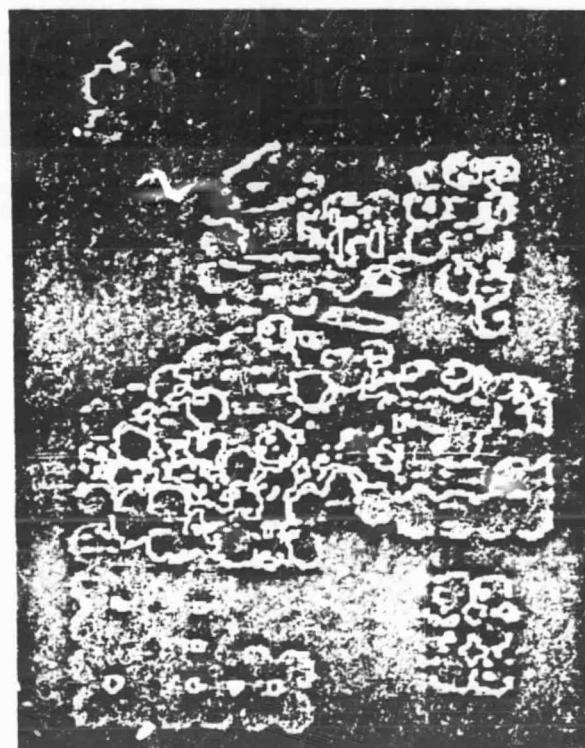


Figure 2. Difference image for agricultural environment.

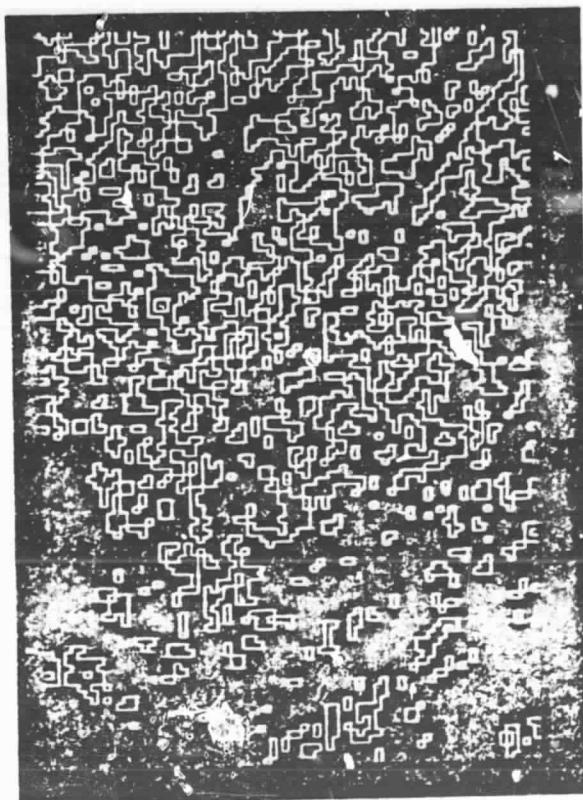


Figure 3. Difference image for range environment.

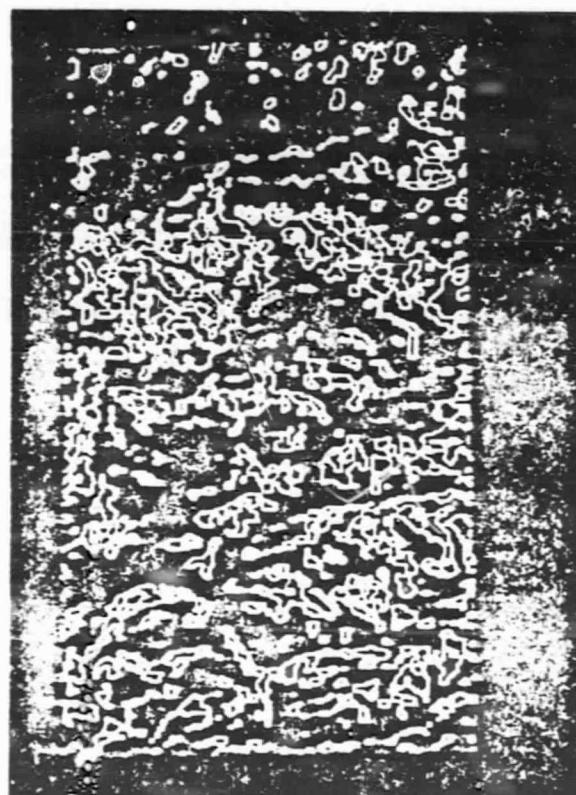


Figure 4. Difference image for forest environment.

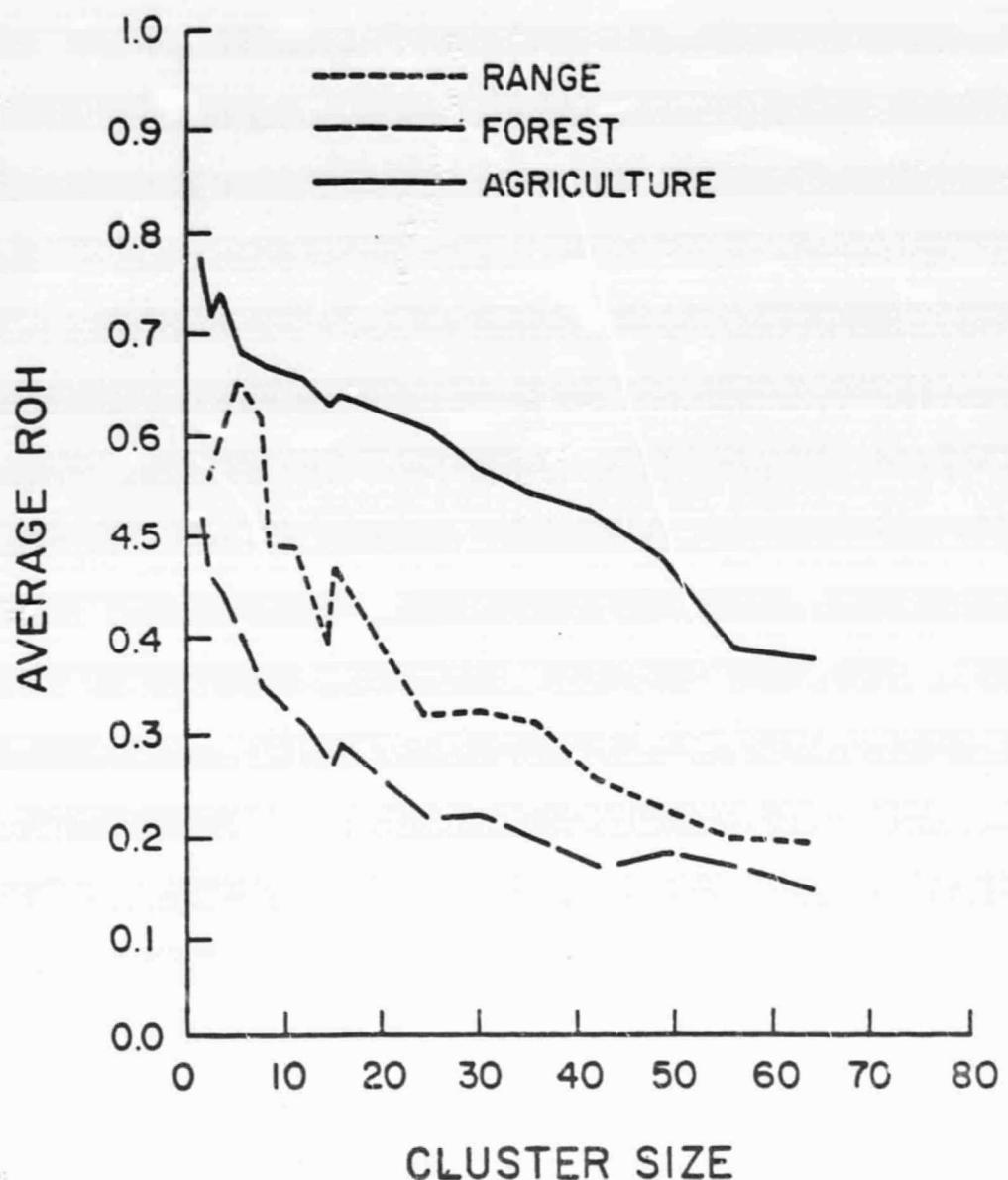


Figure 5. A plot of average ROH vs. cluster size for each vegetation environment.

3.6 Conclusions

As expected the agriculture environment was the most homogeneous because of the large field sizes, while the forest environment was the most heterogeneous. The range environment had a mixture of large areas and small diverse areas and therefore fell somewhere between the agriculture and forest sites. These spatial patterns can be seen by looking at the difference images and also in the plot of ROH vs. cluster size. Remember that a large value of ROH (i.e., close to one) means that the cluster is more homogeneous. Therefore, as seen in Figure 5, the agriculture environment has the largest ROH while the forest site has the smallest.

Also the plot of ROH vs. cluster size dictates some guidelines on what cluster sizes to use. Note that between 0 and 20 pixels/cluster ROH decreases rather quickly while after around 20 pixels/cluster the improvement (i.e., decrease) in ROH occurs more slowly. This result dictates that large cluster sizes may not be gaining more information while costing more time and money to be researched. Therefore, despite the theoretical notion that ROH should be made to go to zero, it is more practical to use reasonable cluster sizes based on this plot and some economic information.

3.7 Further Work

There is a great deal of additional work to be done in sampling simulation. This project has just begun and we hope to accomplish a

great deal more in the next year. Additional sampling schemes need to be investigated and new data sets collected. A possible new data set that contains both a Landsat classification and a reference map is a section of the San Juan National Forest in Colorado. Further investigation in this area can lead to advances in accuracy assessment procedures.

4.0 Literature Cited

Bishop, Y., S. Fienberg, and P. Holland. 1975. *Discrete Multivariate Analysis: Theory and Practice*. MIT Press: Cambridge, Massachusetts. 575 pp.

Congalton, Russell G. 1981. *The Use of Discrete Multivariate Analysis Techniques for the Assessment of Landsat Classification Accuracy*. Unpublished M.S. thesis. Virginia Polytechnic Institute and State University. Department of Forestry. Blacksburg, Virginia 24061. 111 pp.

Congalton, Russell G., Roy A. Mead, Richard G. Oderwald, and Joel T. Heinen. 1981. *Analysis of forest classification accuracy*. AGRISTARS Report. Nationwide Forestry Applications Program. Cooperative Agreement No. 13-1134. 85 pp.

Fienberg, Stephen E. 1970. *An iterative procedure for estimation of contingency tables*. *Annals of Mathematical Statistics*. Vol. 41. No. 3. pp. 907-917.

Fienberg, Stephen E. 1980. *The Analysis of Cross-Classified Categorical Data*. MIT Press: Cambridge, Massachusetts. 198 pp.

Gregg, Tim W. D., Eric W. Barthmaier, Robert E. Aulds, and Robert B. Scott. 1979. *Landsat operational inventory study*. Division of Technical Services. State of Washington. Department of Natural Resources. Olympia, Washington. 28 pp.

Hoffer, Roger M. 1975. *Natural resource mapping in mountainous terrain by computer analysis of ERTS-1 satellite data*. LARS, Purdue University. LARS Research Bulletin 919. LARS Information Note 061575. 124 pp.

Appendix A

FORTRAN COMPUTER PROGRAM KAPPA

KAPPA

/*PAGE 5=35
* KAPPA WAS WRITTEN AND DOCUMENTED BY
* RUSSELL G. CONGALTON
* DPT. OF FORESTRY, VPISU
* JULY 1979
*/

* THIS PROGRAM WAS DESIGNED TO TEST FOR SIMILAR DEGREES OF AGREEMENT
* BETWEEN TWO OR MORE SQUARE ERROR MATRICES
*
* ME = THE NUMBER OF TABLES OR MATRICES TO BE COMPARED
* NR = NUMBER OF ROWS; ALSO THE NUMBER OF COLUMNS SINCE THE
* MATRIX IS SQUARE
* X(I,J) = THE VALUE IN THE MATRIX FOR ROW I AND COLUMN J
*/

REAL KHAT,LCL
DIMENSION X(20,20),SXR(20),SXC(20),S0(20),STAT(20,20)
L=20
M=0
K=1
C
1
2
3
4
5
6

```

7   10 FORMAT(5,12) NR
    C   100 DO 200 I=1,1
9     SXR(1)=0.0
10    SXC(1)=0.0
11    DO 300 J=1,1
12    X(1,J)=0.0
13    300 CUN(1,NR)
14    C

15    READ(5,20) NR
16    20 FORMAT(12)
17    DO 400 I=1,NR
18    READ(5,30) (X(I,J), J=1,NR)
19    30 FORMAT(12(F6.0))
400  CONTINUE
20
21    READ(5,31)
22    31 FORMAT( AAAAAAAAAAAAAAAAAAAAAAAAAA' )
23    WRITE(6,999)
24    999 FORMAT(12)
25    WRITE(6,31)
26    WRITE(6,32)
27    32 FORMAT(12X)
28    WRITE(6,34)
29    34 FORMAT(12X, 'THE ORIGINAL ERROR MATRIX IS' )
30    WRITE(6,35)
31    35 FORMAT(12X, -----' )
32
33    DO 450 I=1,NR
34    WRITE(6,32) (X(I,J), J=1,NR)
35    450 CONTINUE
    C

```

```

56      XN=0.0      I=1.0R
57      DO 500  J=1.0R
58      SXH(1)=SXH(1)+X(1,J)
59      SXC(J)=SXC(J)+X(1,J)
60      CONTINUE
61      XN=XN+SXH(I)
62      CONTINUE
63
64      TH1=0.0
65      TH2=0.0
66      TH3=0.0
67      TH4=0.0
68      DO 700  I=1.0R
69      TH1=TH1+X(I,I)
70      TH2=TH2+SXR(I)*SXC(I)
71      TH3=TH3+X(I,J)*(SXR(I)*SXC(I))
72      DO 800  J=1.0R
73      TH4=TH4+X(I,J)*(SXR(I)*SXC(J))*X2
74      CONTINUE
75      CONTINUE
76      TH1=TH1/XN
77      TH2=TH2/(XN*X2)
78      TH3=TH3/(XN*X3)
79      KHAI(K)=(TH1-TH2)/(1.0-TH2)
80      VARIANCE(K)=(TH1*(1.0-TH2)/(1.0-TH1)*TH2+(2.0*(1.0-TH2)*(2.0*(1.0-TH2)*(2.0*(1.0-TH2)*X2)))/(1.0-TH2)*X2)/(XN-TH2)
81      A33)/((1.0-TH2)*(1.0-TH1)*TH2+(1.0-TH1)*TH2+(1.0-TH1)*TH2+(1.0-TH1)*TH2)
82      SD(K)=SD(VARNC(K))
83      STATE(K)=KHAI(K)/SD(K)
84
85      C THE STEPS THAT FOLLOW CALCULATE THE 95% CONFIDENCE INTERVAL FOR KHAI
86
87      UCL(K)=KHAI(K)+1.96*SD(K)
88      LCL(K)=KHAI(K)-1.96*SD(K)
89      WRITE(6,40) 'LOWER LIMIT',4X,'KHAI',4X,'UPPER LIMIT',4X,'-----'
90      WRITE(6,45) FORMAT(4X,'-----')
91      WRITE(6,50) TET(K),KHAI(K),SD(K)
92      WRITE(6,51) FORMAT(4X,'-----')
93      WRITE(6,52) FORMAT(4X,'-----')
94      WRITE(6,53) FORMAT(4X,'-----')
95      WRITE(6,54) FORMAT(4X,'-----')
96      WRITE(6,55) FORMAT(4X,'-----')
97      WRITE(6,56) FORMAT(4X,'-----')
98      WRITE(6,57) FORMAT(4X,'-----')
99      WRITE(6,58) FORMAT(4X,'-----')
100     WRITE(6,59) FORMAT(4X,'-----')
101     WRITE(6,60) FORMAT(4X,'-----')
102     WRITE(6,61) FORMAT(4X,'-----')
103     WRITE(6,62) FORMAT(4X,'-----')
104     WRITE(6,63) FORMAT(4X,'-----')
105     WRITE(6,64) FORMAT(4X,'-----')
106     WRITE(6,65) FORMAT(4X,'-----')
107     WRITE(6,66) FORMAT(4X,'-----')
108     WRITE(6,67) FORMAT(4X,'-----')
109     WRITE(6,68) FORMAT(4X,'-----')
110     WRITE(6,69) FORMAT(4X,'-----')
111     WRITE(6,70) FORMAT(4X,'-----')
112     WRITE(6,71) FORMAT(4X,'-----')
113     WRITE(6,72) FORMAT(4X,'-----')
114     WRITE(6,73) FORMAT(4X,'-----')
115     WRITE(6,74) FORMAT(4X,'-----')
116     WRITE(6,75) FORMAT(4X,'-----')
117     WRITE(6,76) FORMAT(4X,'-----')
118

```

```

79 FORMAT (5X, 'THE / STATISTIC IS: ', F10.5)
80
81 K=K+1
82 MEM+1
83 IF(M>LT,ME) GO TO 100
84 WRITE(6,900)
85 900 FORMAT(1X, 'SUMMARY TABLE AND COMPARISONS')
86 WRITE(6,910)
87 910 FORMAT(1X, 'XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX')
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
479
480
481
482
483
484
485
486
487
488
489
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
509
510
511
512
513
514
515
516
517
518
519
519
520
521
522
523
524
525
526
527
528
529
529
530
531
532
533
534
535
536
537
538
539
539
540
541
542
543
544
545
546
547
548
549
549
550
551
552
553
554
555
556
557
558
559
559
560
561
562
563
564
565
566
567
568
569
569
570
571
572
573
574
575
576
577
578
579
579
580
581
582
583
584
585
586
587
588
589
589
590
591
592
593
594
595
596
597
598
599
599
600
601
602
603
604
605
606
607
608
609
609
610
611
612
613
614
615
616
617
618
619
619
620
621
622
623
624
625
626
627
628
629
629
630
631
632
633
634
635
636
637
638
639
639
640
641
642
643
644
645
646
647
648
649
649
650
651
652
653
654
655
656
657
658
659
659
660
661
662
663
664
665
666
667
668
669
669
670
671
672
673
674
675
676
677
678
679
679
680
681
682
683
684
685
686
687
688
689
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
709
710
711
712
713
714
715
716
717
718
719
719
720
721
722
723
724
725
726
727
728
729
729
730
731
732
733
734
735
736
737
738
739
739
740
741
742
743
744
745
746
747
748
749
749
750
751
752
753
754
755
756
757
758
759
759
760
761
762
763
764
765
766
767
768
769
769
770
771
772
773
774
775
776
777
778
779
779
780
781
782
783
784
785
786
787
788
789
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
809
810
811
812
813
814
815
816
817
817
818
819
819
820
821
822
823
824
825
826
827
828
829
829
830
831
832
833
834
835
836
837
838
839
839
840
841
842
843
844
845
846
847
848
849
849
850
851
852
853
854
855
856
857
858
859
859
860
861
862
863
864
865
866
867
868
869
869
870
871
872
873
874
875
876
877
878
879
879
880
881
882
883
884
885
886
887
888
889
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
909
910
911
912
913
914
915
916
917
917
918
919
919
920
921
922
923
924
925
926
927
928
929
929
930
931
932
933
934
935
936
937
938
939
939
940
941
942
943
944
945
946
947
948
949
949
950
951
952
953
954
955
956
957
958
959
959
960
961
962
963
964
965
966
967
968
969
969
970
971
972
973
974
975
976
977
978
979
979
980
981
982
983
984
985
986
987
988
989
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1009
1010
1011
1012
1013
1014
1015
1016
1017
1017
1018
1019
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1109
1110
1111
1112
1113
1114
1115
1116
1117
1117
1118
1119
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1139
1140
1141
1142
1143
1144
1145
1146
1147
1147
1148
1149
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1209
1210
1211
1212
1213
1214
1215
1216
1217
1217
1218
1219
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1309
1310
1311
1312
1313
1314
1315
1316
1317
1317
1318
1319
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1409
1410
1411
1412
1413
1414
1415
1416
1417
1417
1418
1419
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1509
1510
1511
1512
1513
1514
1515
1516
1517
1517
1518
1519
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1609
1610
1611
1612
1613
1614
1615
1616
1617
1617
1618
1619
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1709
1710
1711
1712
1713
1714
1715
1716
1717
1717
1718
1719
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1809
1810
1811
1812
1813
1814
1815
1816
1817
1817
1818
1819
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1909
1910
1911
1912
1913
1914
1915
1916
1917
1917
1918
1919
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2009
2010
2011
2012
2013
2014
2015
2016
2017
2017
2018
2019
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2109
2110
2111
2112
2113
2114
2115
2116
2117
2117
2118
2119
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2209
2210
2211
2212
2213
2214
2215
2216
2217
2217
2218
2219
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2309
2310
2311
2312
2313
2314
2315
2316
2317
2317
2318
2319
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2409
2410
2411
2412
2413
2414
2415
2416
2417
2417
2418
2419
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2509
2510
2511
2512
2513
2514
2515
2516
2517
2517
2518
2519
2519
2520
2521
2522
2523
2524
2525
25
```

APPENDIX B

RESULTS OF ACCURACY ASSESSMENT OF THE SAN JUAN
NATIONAL FOREST R2MAP LAND COVER CLASSIFICATION

The κ statistic, κ , was calculated for each of the eight matrices and for contract classification east and west (CCE, CCW), data base classification east and west (DBE, DBW), abbreviated contract classification east and west (ACE, ACW), and abbreviated data base classification east and west (ADE, ADW). The resulting κ 's, variances, and 95% confidence intervals for κ are displayed in Table B.1. The confidence intervals are displayed graphically in Figure B.1. κ was significantly greater than zero ($\alpha = .05$) for all eastern classification, and was not significantly different from zero ($\alpha = .05$) for all western classification.

Classification accuracy as measured by κ was very low for all matrices. The low κ values, however, may not be entirely due to low classification accuracy. The large number of "NO SYMBOL" categories in each matrix and the small sample sizes, particularly in the western matrices, also contribute to the low κ 's. The extent of this contribution, however, cannot be assessed mathematically.

Comparisons were made of the κ 's for contract classification versus data base classification by location (east and west), for eastern classification versus western classification by classification type (contract and data base), and for abbreviated classification versus full classification by location (east and west). The results appear in Table B.2. κ 's for the stated comparisons were not significantly different ($\alpha = .05$) in any instance.

Further accuracy analyses of the matrices can be conducted to evaluate their usefulness for a particular purpose. For example, a weighting scheme can be used to emphasize categories of interest to wildlife managers while reducing the importance of forest and range categories. Another way to accomplish the same end without weights is to lump together categories whose value to a function is minimal. It is very possible that these maps are quite suitable for one function while being inappropriate for another. These types of analyses can be performed quite easily and quickly if the necessary information (weights and/or categories to be lumped) is made available.

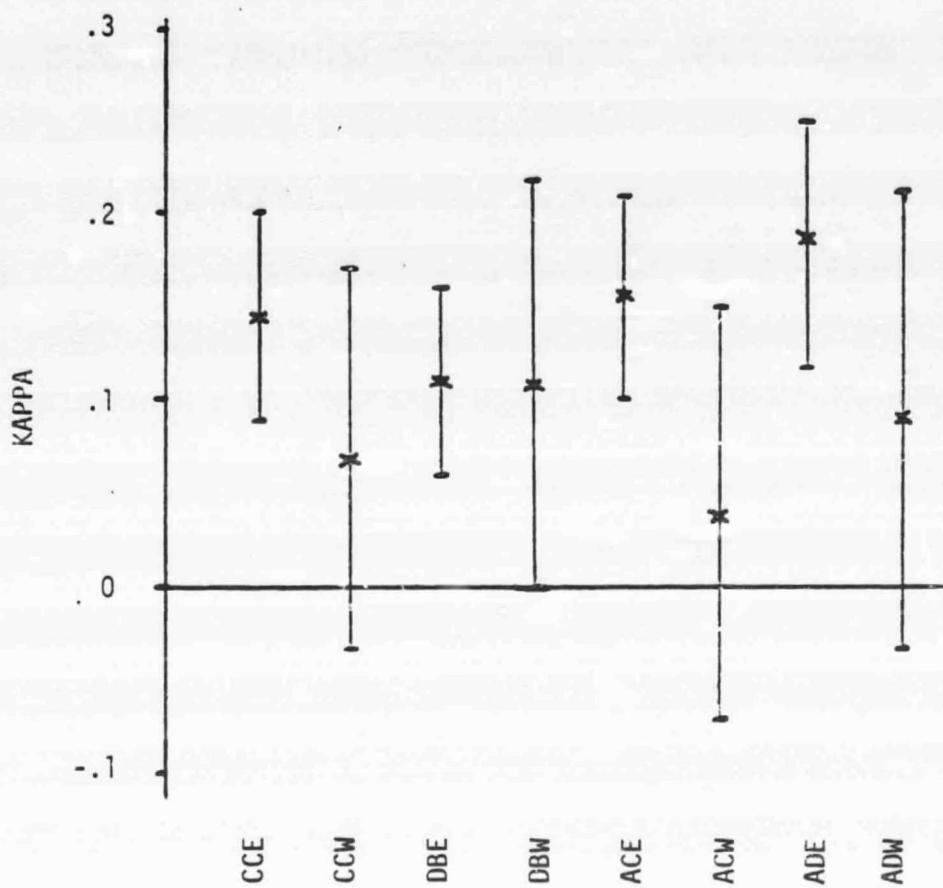
Table B.1. Kappa (κ), variance of Kappa, and 95% confidence interval for Kappa for each classification matrix.

<u>Matrix</u>	<u>κ</u>	<u>σ^2_{κ}</u>	95% Confidence Interval ^{2/} (Lower Limit, Upper Limit)	
CCE	.149	.00076	.094,	.203
CCW	.071	.00281	-.033,	.175
DBE	.109	.00070	.058,	.161
DBW	.108	.00317	-.002,	.218
ACE	.153	.00081	.097,	.208
ACW	.039	.00312	-.071,	.148
ADE	.186	.03260	.122,	.249
ADW	.093	.06179	-.028,	.214

1/ See text for explanation of matrix identification abbreviations.

2/ Confidence interval calculated as $\kappa \pm 1.96\sigma_{\kappa}$.

Figure B.1. Graphical representation of 95% confidence intervals for Kappa for each classification matrix.^{1/}



^{1/} See text for explanation of matrix identification abbreviations.

Table B.2. Tests for significant differences between Kappa's
from selected matrices.^{1/}

<u>Matrices</u>	<u>Z Value</u> ^{2/}
<u>Contract classification versus data base classification by location</u>	
CCE vs. DBE	1.02
CCW vs. DBW	0.48
ACE vs. ADE	0.76
ACW vs. ADW	0.65
<u>Eastern classification versus western classification by classification type</u>	
CCE vs. CCW	1.30
DBE vs. DBW	0.02
ACE vs. ACW	1.82
ADE vs. ADW	1.32
<u>Abbreviated classification versus full classification by location</u>	
CCE vs. ACE	0.11
CCW vs. ACW	0.42
DBE vs. ADE	1.81
DBW vs. ADW	0.18

1/ See text for explanation of matrix identification abbreviations.

2/ Test statistic $Z = \frac{\kappa_A - \kappa_B}{\sqrt{\sigma_{\kappa A}^2 + \sigma_{\kappa B}^2}}$; Z must exceed 1.96 for the κ 's to be significantly different ($\alpha = .05$).

APPENDIX C

ESTIMATING SAMPLE SIZE REQUIREMENTS FOR
ACCURACY ASSESSMENT USING KAPPASample size formula

Sample size calculations for estimating Kappa are based on the confidence interval formula

$$\kappa \pm Z \cdot \sigma_{\kappa}$$

where κ is Kappa, σ_{κ} is the standard deviation of Kappa, and Z is a standard normal deviate. The value of Z may be selected to yield an interval of the desired confidence level.

We will require the estimate of κ to be within $\pm E$ of the true κ , where E is the allowable limit of error. This is equivalent to saying

$$E = Z\sigma_{\kappa}.$$

Since the functional form of σ_{κ} is extremely complex, an approximate, simpler form of σ_{κ} will be used. This form is

$$\sigma_{\kappa} = \sqrt{\frac{p_0(1 - p_0)}{N(1 - p_c)^2}}$$

where p_0 is the actual agreement in the matrix, p_c is the chance agreement in the matrix, and N is the sample size.

The expression for E may now be rewritten as

$$E = Z \sqrt{\frac{p_0(1 - p_0)}{N(1 - p_c)^2}}$$

which leads to

$$N = \frac{p_0(1 - p_0)}{(1 - p_c)^2} \cdot \frac{Z^2}{E^2}$$

With selection of Z and E and estimation of p_0 and p_c this equation may be used to estimate sample size. This equation is equivalent to the statement: Unless a chance error has occurred, the chance being controlled by Z , the estimate of κ will be within $\pm E$ of the true κ .

To implement the sample size equation p_0 and p_c must be estimated. The value p_0 is the proportion of sample observations lying on the main diagonal of the matrix. This fraction may be estimated based on past analysis or an expected result.

The value p_c is calculated as

$$p_c = \sum_{i=1}^t p_{i+} p_{+i}$$

where t is the number of rows (and columns) in the matrix, p_{i+} is the proportion of observations assigned to category i by the classification algorithm, and p_{+i} is the proportion of observations belonging to reference data category i . (The definitions of p_{i+} and p_{+i} may be switched by transposing the matrix.)

The proportion of observations in the i^{th} reference data category, p_{+i} , may be estimated using the assumed proportion of land area in the coverage area that is in category i . While p_{+i} cannot usually be reliably estimated prior to sampling, it is reasonable to assume that adequate classification would result in p_{+i} that is approximately equal to p_{+i} . Therefore, p_c can be rewritten as

$$p_c = \frac{t}{\sum_{i=1}^t p_{+i}^2}$$

These estimates of p_0 and p_c may be combined to estimate σ_k^2 . The given approximation for σ_k^2 has been shown to be generally larger than the true variance, although this will not be true in every case. Therefore the calculated sample size, N , will generally be more than sufficient to attain the desired limit of error.

Example sample size calculation

Unless a 1 in 20 chance occurs, we wish to estimate κ to within ± 1 . The area in question consists of only three distinct categories: Water (category 1); Forest (category 2); Range (category 3). The proportion of the area in each category is estimated to be:

<u>Category No.</u>	<u>Type</u>	<u>Proportion</u>
1	Water	.2
2	Forest	.3
3	Range	.5

In past analyses of the area approximately 60% of the observations fell on the main diagonal of the matrix.

$$\text{Now } Z = 1.96$$

$$E = .1$$

$$p_0 = .6$$

$$p_c = (.2)^2 + (.3)^2 + (.5)^2 = .38.$$

Therefore

$$N = \frac{(.6)(1 - .6)}{(1 - .38)^2} \cdot \frac{(1.96)^2}{(.1)^2} \approx 240$$

APPENDIX D

A PROPOSED METHOD FOR ESTIMATING THE
RELIABILITY OF CHANGE DETECTION

Maps are often used to measure changes in cover types or land use over an interval of time. If the maps produced at each time were perfectly accurate, changes could be known without error. Most maps, however, contain errors in classification that make change detection subject to error. There are two perspectives from which to examine change; the first is a proportion of area basis, the second is a site specific basis. Change from a proportion of area perspective deals with the change in the proportion of an area assigned to a particular category over time. For example, a map made at time 1 identifies 50% of the covered area as water, while a map of the same area made at time 2 identifies 60% of the area as water. No reference to a particular location is made, although the overall results can be applied to a particular location.

Change from a site specific perspective deals with the change of a particular location from one category to another over time. Site specific change detection is extremely important in some map uses, but is more difficult to handle analytically than change in proportion to area. The following discussion will deal only with change from a proportion of area perspective. Hopefully, experience and insight gained in working with proportion of area change can be used to develop methods of dealing with site specific change.

The method outlined below is only a first step in determining the reliability of change detection. Further research is necessary before this method is implemented operationally.

Define two error matrices A and B, where matrix A is produced from mapping an area at time 1, matrix B is produced by mapping the same area at a subsequent time 2, and each matrix is comprised of the same categories. Let p_{ij}^A and p_{ij}^B denote the proportion of sample observations assigned to category i at times 1 and 2, respectively.

Both p_{ij}^A and p_{ij}^B are subject to errors of omission and commission in the classification. If some measure of the reliability of p_{ij}^A and p_{ij}^B could be determined, the reliability of the change over time could be determined.

The agreement measure Kappa, κ , previously defined in this report, provides a type of reliability measure for an entire error matrix. A category specific measure of agreement, similar to Kappa, called κ_i , is defined by Bishop et al., 1975, as

$$\kappa_i = \frac{p_{ii} - p_{i+}p_{+i}}{p_{+i} - p_{i+}p_{+i}}$$

where p_{ij} is the proportion of observations in the i^{th} classification category and i^{th} reference data category, p_{i+} is the proportion of observations in the i^{th} reference data category, and p_{+i} is the proportion of observations in the i^{th} classification category. (The identity of the rows and columns can be exchanged by transposing the matrix.) κ_i has the same characteristics as κ in that it accounts for both chance and actual agreement and has the same range.

If interest is restricted to κ_i such that $0 \leq \kappa_i \leq 1$, then κ_i can serve as a reliability measure of a particular p_{ij+} . It is reasonable to restrict attention to κ_i in this range since negative agreement in an error matrix is an undesirable and hopefully unusual occurrence.

The reliability of the change from p_{ij+}^A to p_{ij+}^B can now be calculated in the same manner as the reliability of a parallel circuit. In the parallel circuit the two components, in this case p_{ij+}^A and p_{ij+}^B , have individual reliabilities, κ_i^A and κ_i^B , and the reliability of the entire circuit is calculated as $\kappa_i^A \cdot \kappa_i^B$. This value can be calculated for each category resulting in a matrix of reliabilities of changes from a category at time 1 to a category at time 2.

The example in Figure D.1. shows the original error matrices A and B at times 1 and 2, respectively, their associated κ_i 's, and the matrix of reliabilities of change. This last matrix can be evaluated cell by cell, or an overall reliability can be found by averaging the cell entries. A further refinement in overall reliability calculation can be achieved by weighting each cell value by importance.

Further research is needed to determine if κ_i can truly be interpreted as a reliability measure, what consequences must be accepted when some κ_i are less than zero, whether this method can be extended to cover unmatched categories between the two maps, and if the method can be used to determine site specific change reliability.

Figure D.1. Example of the proposed methods of determining reliability of change detection.

Error Matrix A

		<u>Reference Data</u>				
		1	2	3	p_{i+}^A	κ_i^A
Classification	1	.31	.03	.02	.36	.75
	2	.02	.20	.05	.27	.61
	3	.04	.05	.28	.37	.68

Error Matrix B

		<u>Reference Data</u>				
		1	2	3	p_{i+}^B	κ_i^B
Classification	1	.14	.02	.05	.21	.54
	2	.03	.26	.02	.31	.73
	3	.05	.04	.39	.48	.71

Reliability Matrix

		κ_1^B	κ_2^B	κ_3^B
		.54	.73	.71
κ_1^A	.75	.41	.55	.53
κ_2^A	.61	.33	.45	.43
κ_3^A	.68	.27	.50	.48

APPENDIX E

Accuracy Assessment of the San Juan National
Forest R2MAP Land Cover Classification

E.1 Introduction

An accuracy assessment was conducted for the San Juan National Forest. Specifically, this included development of error matrices to supplement the preliminary evaluation made by Lockheed Electronics Company,* Inc. Lockheed has used Landsat digital data to map land cover/vegetation for the entire San Juan National Forest from two adjacent scenes according to a classification system agreed to by the Forest Service. Personnel at Virginia Tech worked with managers on the San Juan National Forest to determine the accuracy for (a) the east half of the Forest (from the eastern Landsat scene); (b) the west half of the Forest (from the western Landsat scene); (c) using the classification system agreed to in the Lockheed contract; and (d) according to the classification system used in development of the Forest's "R2MAP" digital data base. An explanation for the two classification systems is given in Appendix F. Also the corresponding R2MAP symbols are given in Tables E.1 and E.2.

E.2 Procedure

The following procedure was used to conduct the accuracy assessment:

* Mazade, A. V., C. A. Underwood, J. F. Ward, and S. S. Yao. 1979. Remote Sensing and Computer-Based Vegetation Mapping in the San Juan National Forest, Colorado. Final Report LEC-13792, Lockheed Electronics Co., Inc. 60 pp.

Table E.1. The contract land cover categories and the corresponding R2MAP symbols.

Land Cover Categories	Symbol	West half of forest	East half of forest
Aspen/Cottonwood	G	95 A4	\$2 A3 94 A4
Aspen/Conifer	BB	No R2MAP Symbol	E2
Ponderosa Pine	A	X5	P1 P3
Spruce-Fir	E	No R2MAP Symbol	S4
Douglas-Fir	C	No R2MAP Symbol	C4 D3
Ponderosa Pine/Oak	Z	F2 F3	No R2MAP Symbol
Conifer/Aspen	AA	No R2MAP Symbol	H4
Oak	K	01 04 02 03 L2	04 03 02
Pinyon/Juniper	M	J4	J1 02 &1
Oak/Conifer	CC	T3	Z5
Rock/Barren	U	Y5 Ø5 /5 A5 V5 =5]5 @5 A5 N5	X5 R5
Willow	O	No R2MAP Symbol]5
Mixed Brush	Q	B4 B3	No R2MAP Symbol
Mesic	R	M5	No R2MAP Symbol
Grass	Y	Z5 G5 <5	G5 =5 /5
Alpine	X	No R2MAP Symbol	M5 :5
Rocky/Grass	DD	6	I5 %5 Ø5
Water	V	W5	W5 #5
Other	W	U5 %5 #5	@5 <5

Table E.2. The data base land cover categories and the corresponding R2MAP symbols.

Land Cover Categories	Symbol	West half of forest	East half of forest
Cottonwood (>30%)	I	No symbol	No symbol
Aspen (>30%)	G	95 A4	\$2 A3 94 A4 #4
Ponderosa Pine (>30%)	A	F3	P1 P3
Spruce-Fir (>30%)	E	No symbol	S4
Douglas-Fir (>30%)	C	No symbol	C4 D3
Oak (>30%)	K	04 03 02 T3	04 03 02
Pinyon Juniper (>30%)	M	J4	No symbol
Cottonwood (10-30%)	J	No symbol]5
Aspen (10-30%)	H	No symbol	E2
Ponderosa Pine (10-30%)	B	X5 F2	Z5
Spruce-Fir (10-30%)	F	No symbol	No symbol
Douglas-Fir (10-30%)	D	No symbol	No symbol
Oak (10-30%)	L	L2 01	No symbol
Pinyon Juniper (10-30%)	N	No symbol	J1 02 &1
Rock/Barren	U	V5 =5]5 @5 A5 N5 Y5 Ø5 R5 /5	X5 R5
Willow	O	No symbol	No symbol
Mixed Brush	P	B4 B3	No symbol
Sage	Q	No symbol	No symbol
Meadow	R	M5	M5
Sonoran	S	G5	Ø5

Table E.2. (Continued)

Land Cover Categories	Symbol	West half of forest	East half of forest
Montane	T	Z5 <5 6	G5 =5
Alpine Xeric	Y	No symbol	I5 %5 :5
Alpine Mesic	X	No symbol	/5 -5
Water	V	W5	W5 #5
Other	W	U5 %5 #5	@5 <5

Step I.

The staff of the San Juan National Forest visited many areas on the ground and related them to the symbols printed on the R2MAPS. This permitted the local resource managers to develop a detailed definition of the Landsat classification categories. Also, the field crews became familiar with characteristics of each category as they appear on color infrared aerial photography. A set of "photo examples" for each category was made for use by all the photo interpreters. This should help assure consistency in the ground reference data collection.

Step II.

Computer printouts which summarize the number of acres of each Landsat category classified on each quad (from R2MAP) was produced. The forest boundary was used to screen only those R2MAP cells within the National Forest. All private lands within the forest were deleted. This permitted the forest to be stratified into the various Landsat categories which were each sampled proportionally. (Note that the classification was sampled and not the ground cover.) After the relative proportions of each strata were determined the number of pixels (3 ac. cells) within each quad that should be sampled by each category were computed.

Step III.

A list of random coordinates were compiled for use in selecting sample cells within each individual quad. Cells were sampled without

replacement until the desired number of cells needed for each category was reached. The location of each cell kept and used in the evaluation were transferred to its corresponding location on the topographic map and assigned a sample number.

Step IV.

Each R2MAP cell selected was next transferred from the topographic map to the 1:30,000 scale 9 x 18 inch aerial photography (flown in September, 1981) and delineated on a transparent overlay fastened to the photo. No indication of how Landsat classified each cell was put on the overlay. (This would bias the photo interpretation.) Only the sample number was next to each cell.

Step V.

Three independent photo interpreters assigned a Landsat category to each sample block according to the category definitions and "photo examples" developed in step I. Complete interpretation agreement among the three interpreters was mandatory. All differences in category assignment were resolved. This required the photo interpreters to meet and "negotiate" a proper interpretation.

Step VI.

Virginia Tech designed a technique and administered a test to assure that consistent photo interpretation was achieved. Also, Virginia Tech

designed the forms for recording all data, reviewed the category definitions and all procedures. Finally, Virginia Tech compiled and reported the final error matrix for the R2MAP accuracy. The data were compiled so that one matrix was produced for the east portion (i.e., east Landsat scene) and one for the west (i.e., west Landsat scene) under each classification system. These matrices are given in Tables E.3, E.4, E.5, E.6, E.7, E.8, E.9, and E.10. Note that there were several instances where there was no R2MAP symbol which corresponds to the land cover categories under both the contract or data base classification systems.

E.3 Sample Size Determination

Map accuracy was determined using the \hat{K} statistic (Bishop, Fienberg, and Holland, 1975):

$$\hat{K} = \frac{N \sum_{i=1}^n x_{ii} - \sum_{i=1}^n x_{i+} x_{+i}}{N^2 - \sum_{i=1}^n x_{i+} x_{+i}}$$

where

N = number of observations,

n = number of categories,

x_{ii} = number of observations classified as category i by both photo interpretation and the Landsat classification algorithm,

x_{i+} = number of observations classified as category i by photo interpretation,

x_{+i} = number of observations classified as category i by the Landsat classification algorithm.

Table E.3. Contrast Classification - East Half.

CLASSIFICATION (R2 MAP SYMBOL)		REFERENCE DATA (PHOTO INTERPRETATION)																				
		G	BB	A	E	C	Z	AA	K	M	CC	U	O	Q	R	Y	X	OO	V	W	TOTAL	
\$2 A3	8	1															1	2			13	
44 94																						
E2	5	2	2	1	1	1	1	1	3								2				18	
P1 P3		1	2	5	2							2	1	1		1	1	1				17
S4			9																		9	
C4 03	3	7	12	26	7	1	6	2								2					67	
NO SYMBOL																					0	
H4	7			5	4				1												18	
04 03	7	1	2	3	2	1	5									8	1				30	
02																						
J1 32		3	3			1	1	1				1	1	1	1	1	1	1			15	
A1																						
Z5																					1	
X5 R5			2								9	2				1		1			15	
I5																					0	
NO SYMBOL																					0	
NO SYMBOL																					0	
G5 +5				1								1			3		1				7	
-5																						
M5 +5						1					1										2	
I5 -5							1					3	2								6	
G5																						
W5 +5			2								1							1			4	
-5															1							
05 +5																					1	
TOTAL	31	12	21	37	17	2	10	13	1	1	16	9	2	2	20	4	2	1	1	223		

Table E.4. Contract Classification - West Half.

Table E.5. Abbreviated Contract Classification -
East Half.

REFERENCE DATA																	
CLASSIFICATION (R2 MAP SYMBOL.)	G	B8	A	E	C	AA	K	M	CC	U	O	Y	X	DD	V	W	
82 A3	8	1			1							2					
A4 94																	
E2	5	2	2	1	1	1	3					2					
P1 P3		1	2	5	2					2	1	1	1	1			
S4			9														
C4 D3	3	7	12	26	7	6	2		1			2					
H4	7		5	4		1						1					
04 03	7	1	2	3	2	1	5					8	1				
02																	
J1 02		3	3		1	1	1		1	1	1	1	1				
B1																	
Z5					1												
X5 R5			2						9	2		1					
J5																	
G5 +5	1	1									1	3	1				
/5																	
M5 .5						1					1						
I5 %5				1					3	2							
.05																	
W5 @5			2						1				1				
@5 <5										1							

Table E.6. Abbreviated Contract Classification - West Half.

Table E.7. Data Base Classification - East Half.

CLASSIFICATION (H2 MAP SYMBOL)	REFERENCE DATA (PHOTO INTERPRETATION)																				TOTAL				
	I	G	4	E	C	K	M	J	H	8	F	D	L	N	U	O	P	Q	R	S	T	Y	X	V	W
NO SYMBOL																								0	
S2 43 54																								0	
H4 34																								31	
53																								161	
																								9	
S4																								0	
33																								67	
C4 03																								29	
02																								0	
NO SYMBOL																								0	
																								0	
15																								0	
E2																								8	
15																								0	
NO SYMBOL																								0	
NO SYMBOL																								0	
NO SYMBOL																								0	
02																								6	
15																								0	
15																								15	
NO SYMBOL																								0	
SYMBOL																								0	
NO SYMBOL																								0	
45																								0	
25																								0	
15																								0	
15																								4	
15																								4	
25																								6	
-5 .5																								4	
-5 .5																								4	
45 45																								4	
45																								1	
45																								1	
TOTAL	0	41	23	61	19	14	1	0	5	3	10	0	2	0	13	10	2	0	0	2	10	1	3	1	323

Table E.8. Data Base Classification - West Half.

Table E.9. Abbreviated Data Base Classification -
East Half.

CLASSIFICATION (R2 MAP SYMBOL)	REFERENCE DATA																		TOTAL
	G	A	E	C	K	J	H	B	N	U	R	S	T	Y	X	V	W		
S2 A3																			31
H4 A4 94	16		5	5	1		1	1					2						
P1 P3	1	2	5	2						1			1						12
S4			9																9
C4 D3	9	14	29	9	3		1						2						67
D4 D3 02	8	2	3	2	5		1						3						24
	15																		0
E2	6	1	3	1	3		2	2											18
Z5			1																1
U1 B1	3	3		1					1	1		1	1						11
X5		2						7					1	1					
W5																			
M5																			
Q5								1											1
G5 +5	1	1								1	1								4
I5 9.5			1					2											3
75										1		1							2
W5			1						1								1		3
@5																			0
45																			
	41	23	62	19	14	0	5	3	0	13		2	10	1	3	1	1	1	198

Table E.10. Abbreviated Data Base Classification - West Half.

CLASSIFICATION (R2 MAP SYMBOL)

REFERENCE DATA														
	G	A	K	M	S	L	U	P	R	S	T	V	W	TOTAL
94 A4														0
F3		2												2
04 03 02														
T3			3										1	4
J4							1							1
X5 F2		2	2										1	5
<2 01			2					1						3
R5 V5 <5]5@5					1									1
/5 A5 N5 Y5 05														
B4 B3			1	4										2
M5														0
G5	1	1	3	2		1								8
Z5 6			1											1
<5														
W5														0
U5 %5 #5														0
TOTAL	1	5	13	3		2	1						2	27

The null hypothesis of \hat{K} equal to zero can be tested statistically using the asymptotic variance. The asymptotic variance of \hat{K} , $\hat{\sigma}^2(\hat{K})$, is available, but was not stated here because of its complexity.

The factors affecting the size of $\hat{\sigma}^2(\hat{K})$ that can be controlled by sampling are the number of observations, N , and the number of categories, n . Since the number of categories is set by map requirements, only the number of observations can be controlled.

A small presample was gathered to provide information about the size of the variance for a given sample size in this particular situation. A final sample size for determining overall map accuracy can then be selected using the presample information as a base.

The sources of agreement and disagreement between classification by photo interpretation and by Landsat algorithm can be investigated using the techniques of categorical data analysis (Bishop, Fienberg, and Holland, 1975). Sample sizes necessary for these techniques are fixed by the analysis method rather than by the degree of precision desired; sample sizes below a certain threshold are simply too low to allow analysis.

The usual sample size required to perform categorical data analysis is five times the square of the number of categories. This sample size is expected to be considerably larger than that required to determine overall map accuracy, but the larger sample size is required if the sources of error in the map are to be identified.

E.4 Summary

The number of samples which were taken (due to personnel time) for the accuracy assessment was too small to give reliable results at any specified precision level. However, it is clear that the R2MAP data (i.e., 3 acre cell category labels) are quite different from the consensus of the three photo interpreters. These errors may in large part be due to misregistration in the Landsat classification. Also, some error could be attributed to the process of resampling the 1 acre Landsat pixels to form the 3 acre R2MAP cell classifications. Analyses of this assessment are given in Appendix B.

The utility of the land cover data in R2MAP for use by the San Juan National Forest will have to be judged by the Forest Service personnel.

APPENDIX F

Explanation of the Two Land Cover Classification
Systems used in the San Juan National Forest
Accuracy Assessment

Attachment 2

SAN JUAN NATIONAL FOREST
REMOTE SENSING PROJECTContract Classification of Cover Classes
for Accuracy Assessment

The following classification system further defines the cover types for which the R2MAP symbols were developed. The original definitions in Exhibit A, p.A-4 of the Remote Sensing and Computer Based Vegetation Mapping in the San Juan National Forest, Colorado. Final Report for USDA/FS Contract 53-82x9-8-2338 October 11, 1978 - September 1, 1979, were followed as closely as possible to maintain as much consistency as possible with the work already completed. The primary problems with the existing definitions was their tendency to overlap. The cover type key presented here will reduce this tendency.

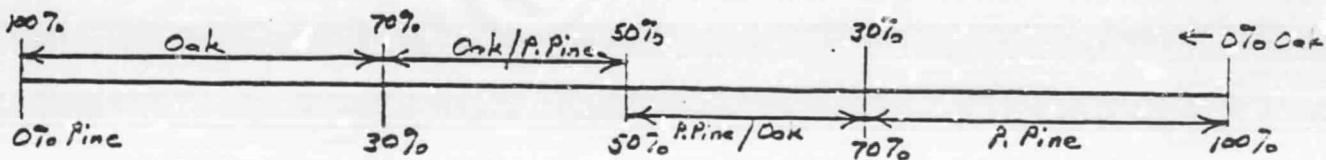
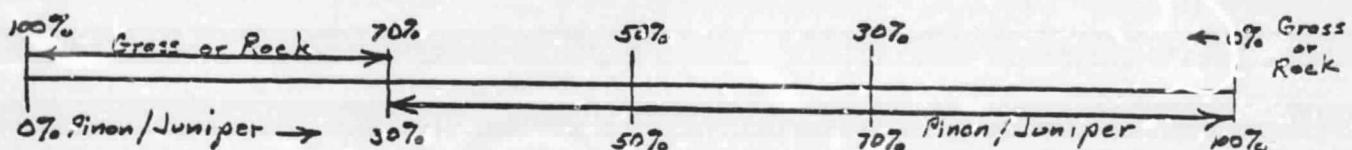
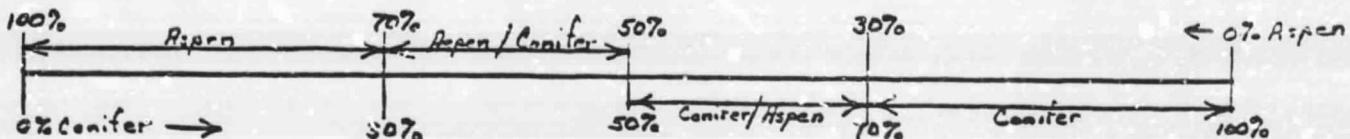
The attached diagram provides a schematic view of the key. These cover classes and cover types apply to 3 acre cells which are the basic unit in R2MAP.

The second attached key is for the Forest data base.

COVER TYPE DIVISIONS

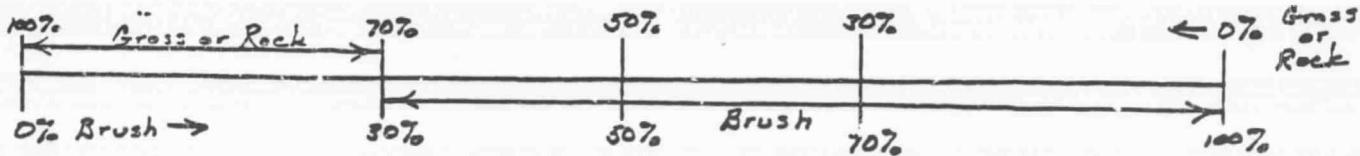
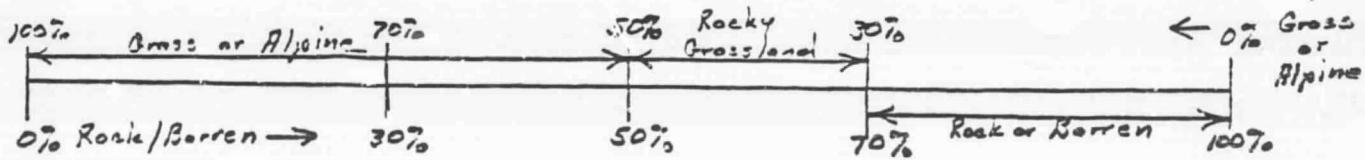
FOREST - Cells with crown cover > 30%

- % refers to % of existing crown or ground cover in a 3 acre cell.



NON FOREST - Cells with < 30%

- % refers to % of the cell not covered by tree crown cover.



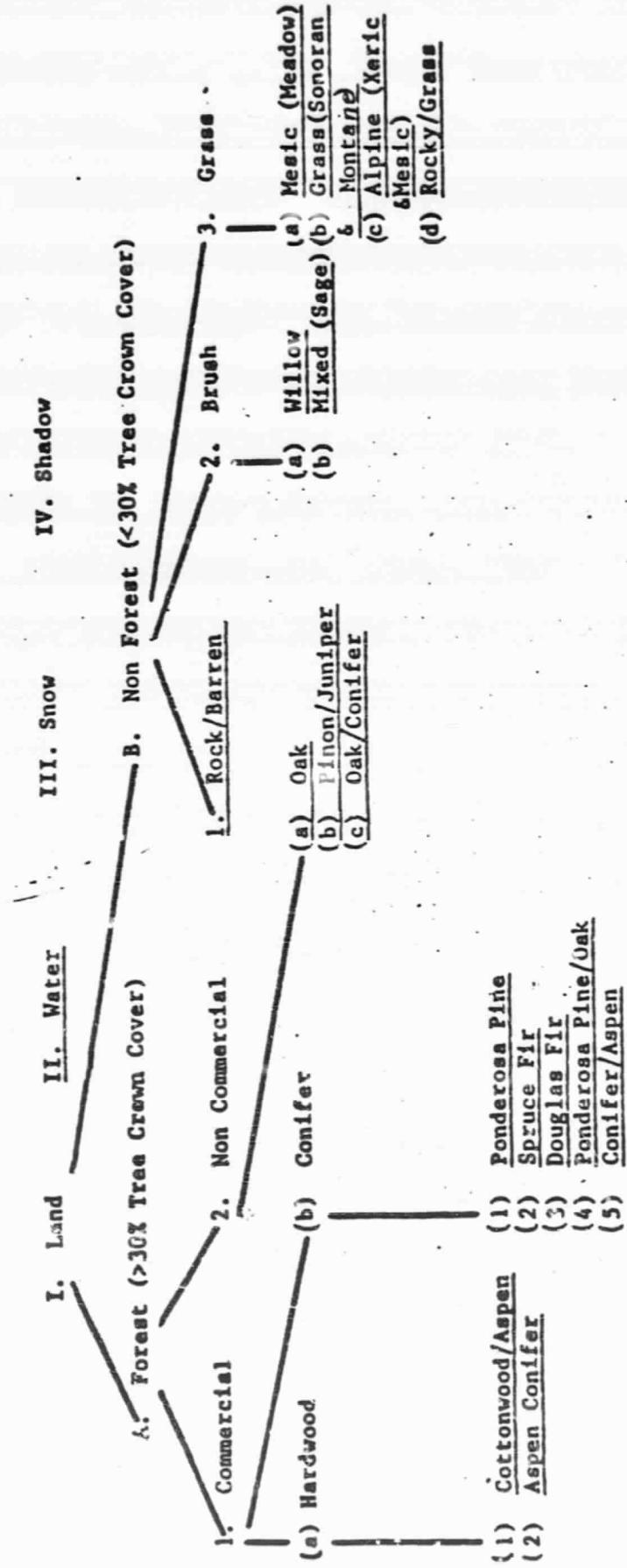
Note: Within the forested cover types, species dominance drives the classification system. Between forest and nonforest cover types and within the nonforest types a hierarchical system exists. Forest types override nonforest types. Brush overrides grass types.

SAM JUAN NATIONAL FOREST
 Ground Cover Type Key-Contract
 (Cover Types are underlined)

I. Lands	IB. Non Forest
IA. Forest	IB1. <u>Rock Barren</u>
IA1. Commercial	IB2. Brush
IA1(a). Hardwood	IB2(a). <u>Willow</u>
IA1(a)(1). <u>Aspen/Cottonwood</u>	IB2(b). <u>Mixed Brush (Sage)</u>
IA1(a)(2). <u>Aspen/Conifer</u>	IB3. Grass
IA1(b). Conifer	IB3(a). <u>Mesic (Meadows)</u>
IA1(b)(1). <u>Ponderosa Pine</u>	IB3(b). <u>Grass (Sonoran & Montane)</u>
IA1(b)(2). <u>Spruce-Fir</u>	IB3(c). <u>Alpine (Xeric & Mesic)</u>
IA1(b)(3). <u>Douglas-fir</u>	IB3(d). <u>Rocky/Grass</u>
IA1(b)(4). <u>Ponderosa Pine/Oak</u>	
IA1(b)(5). <u>Conifer/Aspen</u>	
IA2. Non Commercial Forest	II. <u>Water</u>
IA2(a). <u>Oak</u>	III. Snow
IA2(b). <u>Pinon/Juniper</u>	IV. Shadow
IA2(c). <u>Oak/Conifer</u>	

SAN JUAN NATIONAL FOREST
 Ground Cover Type Key - Contract
 (Cover types are underlined)

72



San Juan National Forest
Ground Cover Type Key - Contract

The objective of this key is to classify the total area within the San Juan National Forest into one of 18 cover types by three acre cells.

I. LAND - Cells covered by > 50% land.

Yes - Go to I.A.
No - Go to II.

I. A. FOREST - Cells covered by tree species >30% crown cover. (Trees are further defined as woody vegetation capable of producing a woody stem ≥ 12 feet in height. This includes oak and other tree species that ≤ 12 feet in height due to site limiting conditions.)

Yes - Go to I.A. 1. or 2.
No - Go to I.B.

I. A.1. COMMERCIAL FOREST - >50% of the crown cover is one or more of the following commercial species: ponderosa pine, spruce-fir, Douglas-fir, aspen, or cottonwood.

Yes - Go to I.A. 1. (a) or (b)
No - Go to I.A. 2.

I. A.1. (a) HARDWOOD - >50% of the crown cover is one or more of the following hardwood species: Aspen or Cottonwood.

Yes - Go to I.A. 1. (a) (1) or (2)
No - Go to I.A. 1. (b)

I. A.1. (a) (1) ASPEN/COTTONWOOD - >70% of the crown cover is Aspen or Cottonwood.

Yes - The cover type is Aspen/Cottonwood.
No - Go to I.A. 1. (a) (2).

I. A.1. (a) (2) ASPEN/CONIFER - Aspen crown cover is >50% but not >70%. The conifer crown cover is <50% but not <30%.

Yes - The cover type is Aspen/Conifer.

I. A.1. (b) CONIFER - >50% of the crown cover is one or more of the following conifer species: ponderosa pine, spruce-fir, or Douglas-fir.

Yes - Go to I.A.1. (b) (1), (2), (3), (4) or (5).

I. A.1. (b) (1) PONDEROSA PINE - >70% of the crown cover is ponderosa pine.

Yes - The cover type is ponderosa pine.

No - Go to I.A.1. (b) (2), (3), (4), or (5).

I. A.1. (b) (2) SPRUCE-FIR - >70% of the crown cover is mixed spruce and fir.

Yes - The cover type is spruce-fir.

No - Go to I.A. 1. (b) (3), (4) or (5).

I. A.1. (b) (3) DOUGLAS-FIR - >70% of the crown cover is mixed Douglas-fir and white fir.

Yes - The cover type is Douglas-fir.

No - Go to I. A. 1. (4) or (5).

I. A.1. (b) (4) PONDEROSA PINE/OAK - Ponderosa pine crown cover is >50% but not >70%. The oak crown cover is < 50% but not < 30%.

Yes - The cover type is ponderosa pine/oak.

No - Go to I.A.1. (b) (5).

I. A.1. (b) (5) CONIFER/ASPEN - Conifer crown cover is > 50% but not > 70%. The aspen crown cover is < 50% but not < 30%.

Yes - The cover type is Conifer/Aspen.

I. A.2. NONCOMMERCIAL FOREST - >50% of the crown cover is one or more of the following noncommercial species: pinon pine, juniper, or oak.

Yes - Go to I.A. 2. (a) (b) or (c)

I. A.2. (a) OAK - >70% of the crown cover present is oak.

Yes - The ground cover type is oak.

No - Go to I.A. 2. (b) or (c).

I. A.2. (b) PINON/JUNIPER - >50% of the crown cover present is pinon/juniper.

Yes - The ground cover type is pinon/juniper.

No - Go to I.A.2. (c).

I. A.2. (c) OAK/CONIFER - Oak crown cover is >50%, but not >70%. The conifer crown cover is < 50% but not < 30%.

Yes - The ground cover type is oak/conifer.

I. B. NONFOREST - Cells covered by < 30% crown cover of tree species.
Yes - Go to I.B., 1. 2. or 3.

I. B.1. ROCK/BARREN - < 30% vegetative ground cover is present.

Yes - Cover type is Rock/Barren.

No - Go to I.B. 2 or 3.

I. B.2. BRUSH - >30% of the area is covered by brush species.

Yes - Go to I.B. 2. (a) or (b)

No - Go to I.B. 3.

I. B.2. (a) WILLOW (Brush) - >30% willow crown cover is present.

Yes - The cover type is Willow.

No - Go to I.B. 2. (b).

I. B.2. (b) MIXED BRUSH (Sage) - >30% of the area is brush other than oakbrush or willows.

Yes - The cover type is Mixed Brush (Sage).

I. B.3. GRASS - >30% of the area is grass and herbaceous plants.

Yes - Go to I.B. 3. (a) (b) (c) or (d).

I. B.3. (a) WET or MESIC GRASSLAND - The area is dominated by grasses and other herbaceous plants requiring constant water availability. The elevation range for this cover type is 6,500 feet to 11,000 feet.

Yes - The cover type is Wet or Mesic Grassland.

No - Go to I. B. 3. (b), (c) or (d).

I. B.3. (b) GRASSLAND (Sonoran and Mountain) - The area is dominated by grasses and other herbaceous plants. The elevation range is 6,500 feet to 11,000 feet.

Yes - The cover type is Grassland.

No - Go to I.B.3. (c) or (d).

I. B.3. (c) ALPINE (Xeric and Mesic) - The area is above timberline and dominated by grass and other herbaceous plants. The elevation range of this cover type is > 11,000 feet.

Yes - The cover type is ALPINE (Xeric and Mesic).

No - Go to I.B. 3. (d).

I. B.3. (d) ROCKY/GRASSLAND - The area is dominated by grasses and other herbaceous species. Rock and barren soil cover >30% but <50% of the area. The elevation range of the area is >6,500.

Yes - The cover type is Rocky/Grassland.

II. WATER - Cells covered by >50% water.

Yes - Ground cover type is water.

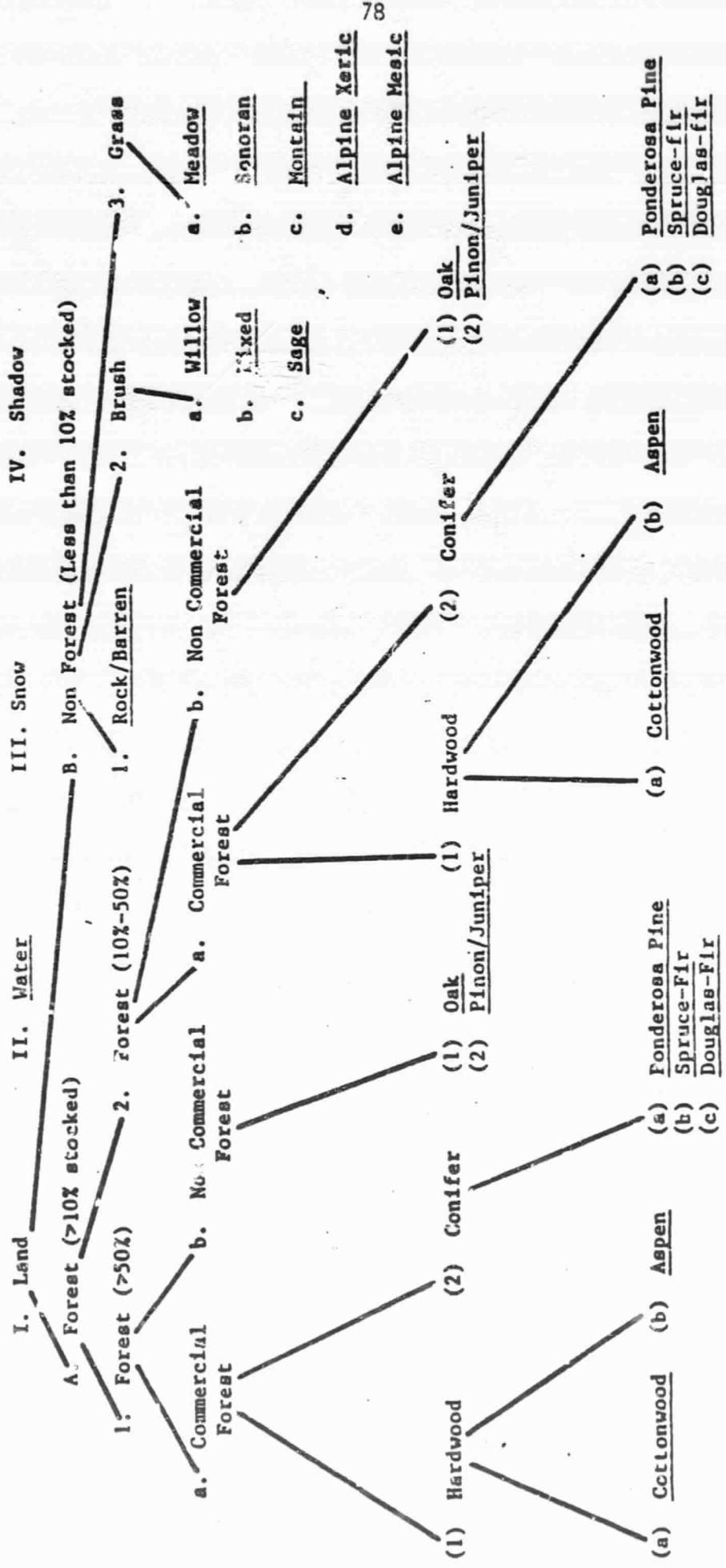
III. SNOW - This cover class division is included because areas will seasonally be covered with snow. Landsat will record this information, if present. Snow is not a valid ground cover type.

IV. SHADOW - This cover class division is included because steep topography on the Forest produces shaded cells regardless of sun angle. Landsat will record this information, if present. Shadow is not a valid ground cover type.

San Juan National Forest
 Ground Cover Type Key - Data Base
 (Cover types are underlined.)

I.	Lands
I.A.	Forest (> 10% stocked)
I.A.1.	Forest (> 30%)
I.A.1.a.	Commercial Forest
I.A.1.a.(1)	Hardwood
I.A.1.a.(1)(a)	<u>Cottonwood</u> (> 30%)
I.A.1.a.(1)(b)	<u>Aspen</u> (> 30%)
I.A.1.a.(2)	Conifer
I.A.1.a.(2)(a)	<u>Ponderosa Pine</u> (> 30%)
I.A.1.a.(2)(b)	<u>Spruce-fir</u> (> 30%)
I.A.1.a.(2)(c)	<u>Douglas-fir</u> (> 30%)
I.A.1.b.	Non-Commercial
I.A.1.b.(1)	<u>Oak</u> (> 30%)
I.A.1.b.(2)	<u>Pinon Juniper</u> (> 30%)
I.A.2.	Forest (10-30%)
I.A.2.a.	Commercial Forest
I.A.2.a.(1)	Hardwood
I.A.2.a.(1)(a)	<u>Cottonwood</u> (10-30%)
I.A.2.a.(1)(b)	<u>Aspen</u> (10-30%)
I.A.2.a.(2)	Conifer
I.A.2.a.(2)(a)	<u>Ponderosa Pine</u> (10-30%)
I.A.2.a.(2)(b)	<u>Spruce-fir</u> (10-30%)
I.A.2.a.(2)(c)	<u>Douglas-fir</u> (10-30%)
I.A.2.b.	Non-Commercial
I.A.2.b.(1)	<u>Oak</u> (10-30%)
I.A.2.b.(2)	<u>Pinon Juniper</u> (10-30%)
I.B.	Non-Forest (less than 10% stock)
I.B.1.	<u>Rock/Barren</u>
I.B.2.	Brush
I.B.2.a.	<u>Willow</u>
I.B.2.b.	<u>Mixed Brush</u>
I.B.2.c.	<u>Sage</u>
I.B.3.	Grass
I.B.3.a.	<u>Meadow</u>
I.B.3.b.	<u>Sonoran</u>
I.B.3.c.	<u>Montane</u>
I.B.3.d.	<u>Alpine Xeric</u>
I.B.3.e.	<u>Alpine Mes'c</u>
II.	<u>Water</u>
III.	Snow
IV.	Shadow

SAN JUAN NATIONAL FOREST
 Ground Cover Type Key - Data Base
 (Cover types are underlined)



San Juan National Forest
Ground Cover Type Key - Data Base

The objective of this key is to classify the total area within the San Juan National Forest into one of 25 cover types by three acre cells.

I. LAND - Cell covered by >50% land.

Yes - Go to I.A.
No - Go to II.

I. A. FOREST (>10% stocked) - Cells covered by tree species >10% crown cover. (Trees are further defined as woody vegetation capable of producing a woody stem \geq 12 feet in height. This includes oak and other tree species that are < 12 feet in height due to site limiting conditions.)

Yes - Go to I.A.1. or 2.
No. - Go to I.B.

I. A.1. FOREST (>30%) - Cells covered by tree species >30% crown cover.

Yes - Go to I.A.1. a. or b.
No - Go to I.A.2.

I.A.1.a COMMERCIAL FOREST - >50% of the crown cover is one or more of the following commercial species: ponderosa pine, spruce-fir, Douglas-fir, aspen, or cottonwood.

Yes - Go to I.A.1.a.(1) or (2).
No - Go to I.A.1.b.

I. A.1.a.(1) HARDWOOD - >50% of the crown cover is one or more of the following hardwood species: Aspen or Cottonwood.

Yes - Go to I.A.1.a.(1) (a) or (b).
No - Go to I.A.1.a.(2).

I. A.1.a.(1)(a) COTTONWOOD - >50% of the crown cover is Cottonwood.

Yes - The cover type is Cottonwood (>30%).
No - Go to I.A.1.a.(1).(b).

I.A.1.a.(1)(b) ASPEN - >50% of the crown cover is aspen.

Yes - The cover type is aspen (>30%).

I.A.1.a.(2) CONIFER - >50% of the crown cover is one or more of the following conifer species: ponderosa pine, spruce-fir, or Douglas-fir.

Yes - Go to I.A.1.a.(2) (a), (b), or (c).

I.A.1.a.(2)(a) PONDEROSA PINE - >50% of the crown cover is Ponderosa Pine.
Yes - The cover type is Ponderosa Pine (>30%).
No - Go to I.A.1.a.(2) (b) or (c).

I.A.1.a.(2)(b) SPRUCE-FIR - >50% of the crown cover is spruce-fir.
Yes - The cover type is spruce-fir (>30%).
No - Go to I.A.1.a.(2)(c).

I.A.1.a.(2)(c) DOUGLAS FIR - >50% of the crown cover is mixed Douglas fir and white fir.
Yes - The cover type is Douglas fir (>30%).

I.A.1.b. NON-COMMERCIAL - >50% of the crown cover is one or more of the following non-commercial species: pinon pine, juniper, or oak.
Yes - Go to I.A.1.b. (1) or (2).

I.A.1.b.(1) OAK - >50% of the crown cover is oak.
Yes - The cover type is oak (>30%).
No - Go to I.A.1.b.(2).

I.A.1.b.(2) PINON-JUNIPER - >50% of the crown cover is pinon/juniper.
Yes - The cover type is pinon-juniper (>30%).

I.A.2. FOREST (10-30%) - Cells covered by tree species. 10-30% crown cover.
Yes - Go to I.A.2. a. or b.

I.A.2.a. COMMERCIAL FOREST ->50% of the crown cover is one or more of the following commercial species: ponderosa pine, spruce-fir, Douglas fir, aspen, or cottonwood.
Yes - Go to I.A.2.a. (1) or (2).
No - Go to I.A.2.b.

I.A.2.a.(1) HARDWOOD ->50% of the crown cover is one or more of the following hardwood species: Aspen or cottonwood.
Yes - Go to I.A.2.a.(1) (a) or (b).
No - Go to I.A.2.a.(2).

I.A.2.a.(1)(a) COTTONWOOD - >50% of the crown cover is cottonwood.
Yes - The cover type is cottonwood (10-30%).
No - Go to I.A.2.a.(1)(b).

I.A.2.a.(1)(b) ASPEN - >50% of the crown cover is aspen.
 Yes - The cover type is aspen (10-30%).

I.A.2.a.(2) CONIFER - >50% of the crown cover is one or more of the following conifer species: ponderosa pine, spruce-fir, or Douglas fir.
 Yes - Go to I.A.2.a.(2) (a), (b), or (c).

I.A.2.a.(2)(a) PONDEROSA PINE - >50% of the crown cover is ponderosa pine.
 Yes - The cover type is ponderosa pine (10-30%).
 No - Go to I.A.2.a.(2) (b) or (c).

I.A.2.a.(2)(b) SPRUCE-FIR - >50% of the crown cover is spruce-fir.
 Yes - The cover type is spruce-fir (10-30%).
 No - Go to I.A.2.a.(2)(c).

I.A.2.a.(2)(c) DOUGLAS FIR - >50% of the crown cover is mixed Douglas fir and white fir.
 Yes - The cover type is Douglas fir (10-30%).

I.A.2.b. NON-COMMERCIAL - >50% of the crown cover is one or more of the following non-commercial species: pinon pine, juniper, or oak.
 Yes - Go to I.A.2.b. (1) or (2).

I.A.2.b.(1) OAK - >50% crown cover is oak.
 Yes - The cover type is oak (10-30%).
 No - Go to I.A.2.b.(2).

I.A.2.b.(2) PINON-JUNIPER - >50% of the crown cover is pinon juniper.
 Yes - The cover type is pinon-juniper (10-30%).

I.B. NON-FOREST (less than 10% stocking) - Cells covered by <10% crown cover of tree species.
 Yes - Go to I.B., 1., 2., or 3.

I.B.1. ROCK/BARREN - <30% vegetative ground cover is present.
 Yes - Cover type is rock/barren.
 No - Go to I.B. 2 or 3.

I.B.2. BRUSH - >30% of the area is covered by brush species.
 Yes - Go to I.B.2. a, b, or c.
 No - Go to I.B.3.

I.B.2.a. WILLOW (brush) - >30% willow crown cover is present.

Yes - The cover type is willow.

No - Go to I.B.2. b or c.

I.B.2.b. MIXED BRUSH - >30% of the area is brush other than oakbrush, willow, or sage.

Yes - The cover type is mixed brush.

No - Go to I.B.2.c.

I.B.2.c. SAGE BRUSH - >30% sage brush crown cover is present.

Yes - The cover type is sage brush.

I.B.3. GRASS - >30% of the area is grass and herbaceous plants.

Yes - Go to I.B.3. a, b, c, d, or e.

I.B.3.a. MEADOWS (wet) - The area is dominated by grasses and other herbaceous plants requiring constant water availability. The elevation range for this cover type is > 5,500 feet.

Yes - The cover type is meadow.

No - Go to I.B.3. b, c, d, or e.

I.B.3.b. SONORAN GRASSLAND - The area is dominated by grasses and other herbaceous plants. The elevation range is 5,500 to 7,000 feet.

Yes - The cover type is Sonoran grassland.

No - Go to I.B.3. c, d, or e.

I.B.3.c. MONTANE GRASSLAND - The area is dominated by grasses and other herbaceous plants. The elevation range is 6,900 to 9,000 feet.

Yes - The cover type is Montane grassland.

No - Go to I.B.3. d or e.

I.B.3.d. ALPINE XERIC - The area is above timberline and dominated by dry site grasses and other herbaceous plants. The elevation range of this cover type is > 11,000 feet.

Yes - The cover type is Alpine Xeric.

No - Go to I.B.3.e.

I.B.3.e. ALPINE MESIC - The area is above timberline and dominated by wet or moist site grasses and other herbaceous plants. The elevation range of this cover type is > 11,000 feet.

Yes - The cover type is Alpine Mesic.

II. WATER - Cells covered by > 50% water.

Yes - Ground cover type is water.

III. SNOW - This cover class division is included because areas will seasonally be covered with snow. Landsat will record this information, if present. Snow is not a valid ground cover type.

IV. SHADOW - This cover class division is included because steep topography on the Forest produces shaded cells regardless of sun angle. Landsat will record this information, if present. Shadow is not a valid ground cover type.